AD-A221 206

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS NONE
2a. SECURITY CLASSIFICATION AUTHORITY 2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	3. DISTRIBUTION / AVAILABILITY OF REPORT APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFIT/CI/CIA- 90-002D
6a. NAME OF PERFORMING ORGANIZATION AFIT STUDENT AT Pennsylvania State University 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION AFIT/CIA
6c. ADDRESS (City, State, and ZIP Code)	7b. ADDRESS (City, State, and ZIP Code) Wright-Patterson AFB OH 45433-6583
8a. NAME OF FUNDING / SPONSORING ORGANIZATION 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification) (UNCLASSIFIED)
Knowledge of Results in Visual Inspection: A Model of Decision Making

12. PERSONAL AUTHOR(S)
John Micalizzi

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
~~THESIS~~/DISSERTATION	FROM ________ TO ______	1990	168

16. SUPPLEMENTARY NOTATION
APPROVED FOR PUBLIC RELEASE IAW AFR 190-1
ERNEST A. HAYGOOD, 1st Lt, USAF
Executive Officer, Civilian Institution Programs

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

DTIC
ELECTE
MAY 08 1990
S E D

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL ERNEST A. HAYGOOD, 1st Lt, USAF	22b. TELEPHONE (Include Area Code) (513) 255-2259 22c. OFFICE SYMBOL AFIT/CI

DD Form 1473, JUN 86 Previous editions are obsolete. SECURITY CLASSIFICATION OF THIS PAGE

AFIT/CI "OVERPRINT"

The Pennsylvania State University

The Graduate School

KNOWLEDGE OF RESULTS IN VISUAL INSPECTION: A MODEL OF DECISION MAKING

A Thesis in

Industrial and Management Systems Engineering

by

John Micalizzi

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

May 1990

Accession For
NTIS GRA&I X
DTIC TAB
Unannounced
Justification
By
Distribution
Availability
Dist
A-1

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

This research was supported under a National Science Foundation Engineering Research Initiation Grant, No. DMC-8707650, made to J.H. Goldberg. The author is grateful for helpful comments by Dr. J.H. Goldberg and J. Gordon Davy of Westinghouse, Inc., in the initiation of this project.

Chapter 1

INTRODUCTION

The training of industrial workers will be one of the most challenging issues facing American industry during the coming years. There is a growing need to provide effective job training for workers at all levels of technology. Unfortunately, most corporate training programs have neglected to incorporate existing cognitive theory into an overall training strategy. Furthermore, as increasing automation reduces the number of pure manual tasks, cognitive-based tasks such as system monitoring and decision making will become more common.

Problem Statement

One of the biggest problems facing training researchers today is the tendency for high level technology to suddenly appear in the workplace ahead of any overall plan for worker training. In particular, technology such as automation, robotics, and artificial intelligence has been here for years without adequately addressing the issue of operator training. While many of these systems are relatively new and evolving, even traditional tasks such as visual inspection can benefit from a more "cognitive" based approach to training. Such an approach focuses on the internal mental processes of human learning rather than only on the output recorded in performance measures. The term, "cognitive skill component," will be used to describe those mental units of human information processing that can be experimentally manipulated to change task performance.

Previous research on training human monitoring behavior concentrated primarily on the performance benefits of knowledge of results (KR) and stimulus cueing (Adams and Humes, 1963; Colquhoun, 1975). While spurring a large volume of research in these two areas, these early studies could not accurately "model" monitoring performance beyond the conditions in the original experiment. A model must have predictive power for future performance under conditions not explicitly stated beforehand. The strength of a model lies in its generalizability to new situations which increases one's confidence in understanding the underlying components of behavior. A cognitive model is vital to understanding and training human monitoring behavior.

Since the "actions" in human monitoring are primarily covert, and sensitive research studies difficult to design, there is a tendency to view monitoring tasks as inherently low workload, requiring simple yes/no decisions, and easily trainable through repetition. The workload issue is especially deceiving since, although the number of "signals" (e.g., product defects) is usually low, it may be the number of opportunities for a signal that drives the workload demands of a task. Additional experimental evidence is required before any of these "assumptions" concerning human monitoring behavior should be allowed to influence training decisions.

Visual inspection is a special type of system monitoring which has been an important part of the industrial work environment for many years. Only relatively recently, however, have the underlying learning processes involved been studied. Wang and Drury (1987) made one of the first attempts to evaluate the mental demands of an inspection task. Their method, which involved evaluating the relationship between pretested cognitive factors and inspection performance, identified the specific attributes of "attention" and "judgement" as important factors in the search and decision components of inspection. In modeling inspection behavior, these attributes provided a high level description of the underlying mental processes used in inspection. Further refinement of the model will begin to identify the lower level components which can be manipulated in training. Their work, together with others (e.g. Embrey, 1979), has created a long-needed research interest in cognitive based training for inspection.

Background

Visual Inspection Performance

Product inspection is one of the most critical areas to consider for improving industrial quality control. Although consumers may demand defect-free products (Moll, 1976), perfect performance is not possible with human inspectors (Drury, 1982). Automation can eliminate the motivational and bias problems of human inspection, but it cannot exceed the superior decision-making capabilities of the human observer across a wide range of targets. The large trained workforce available also insures that human inspection remains an integral part of any future quality control program.

The most cost effective improvements in human inspection performance logically stem from modifications in training. Since most inspection tasks consist of search followed by decision making, the only way to isolate decision making during training is to eliminate the search requirement. The task learning literature has discussed training for inspection (Embrey, 1975, 1979; Czaja and Drury, 1981), but these studies have emphasized only the overall strategies for training inspectors, not the cognitive skill components needed to develop appropriate training techniques. A basic understanding of these components is essential for developing a useful model of inspector decision-making behavior.

There is general agreement among both quality assurance people and behavioral scientists concerning the structural aspects of inspection. Harris and Chaney (1969) identified the basic elements of visual inspection as interpretation, comparison, decision making, and action. A more detailed model of the perceptual-decision processes occurring during a series of inspections is illustrated in Figure 1 (Adams, 1975). In this model, inspector decision making is based on, among other things, perceived defect probabilities and payoffs stored in memory. While describing inspector decision making at a very general level, this model also represents some of the cognitive factors which affect inspector performance. Specific training manipulations can now be defined and tested to improve inspector performance.

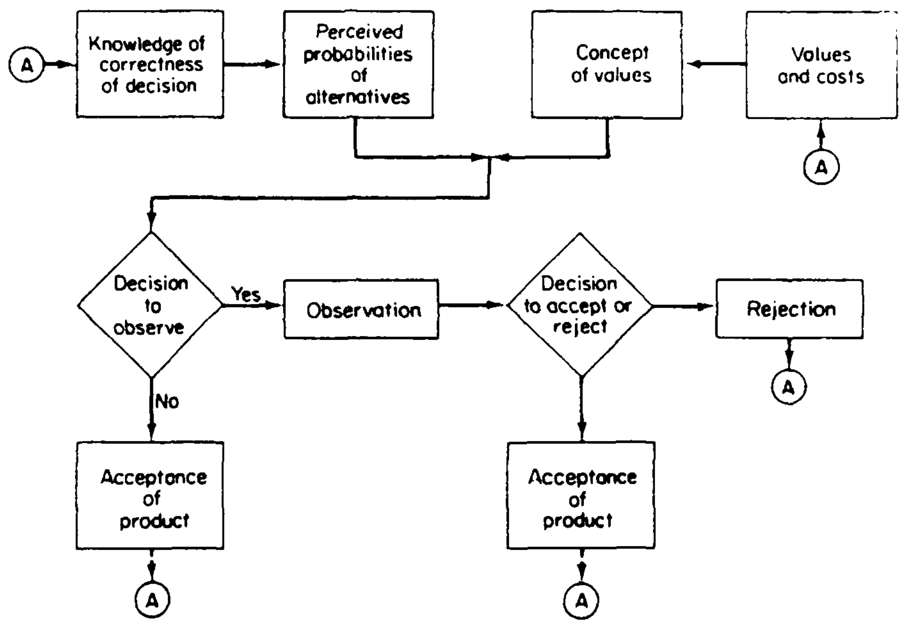


Figure 1. Perceptual-Decision Processes Occurring During Inspection. Source: Adams, (1975)

Despite such models, there is still relatively little understanding of the cognitive abilities that make a good industrial inspector. Research has shown that there are large differences among people in their ability to perform visual inspection tasks. In a study of machined parts inspection performance (Harris and Cheney, 1969), the best inspector observed detected four times as many sample defects as the poorest inspector. Personnel selection tests try to compensate for this disparity by trying to provide the best match possible between a potential worker and a given job (Harris, 1966). Since it is not possible to be certain of pre-selecting the best workers, training is required to bring job performance up to some criterion level. Learning on the job is one way to train industrial inspectors; however, such an approach is reasonable only if there is a good chance to learn from experience. Time on the job alone is not a good predictor of performance. For example, measures of inspection performance obtained under controlled conditions showed no differences in defect detection for inspectors with only 2 months experience compared to inspectors with 48 months of experience (Thresh and Frerichs, 1966). In addition, the effectiveness of industrial inspectors is often exaggerated. It is common for inspection performance to range from fewer than 30% mean defects detected for complex items to no more than 80% for the simplest ones (Harris, 1966). Clearly, visual inspection requires a systematic approach to skill learning similar to other formal training programs.

The effectiveness of training can be evaluated by focusing on several aspects of the learning environment, although performance measures are by far the most dominant. Since the ultimate objective of training is performance improvement, it is not unusual to monitor training progress through measurable changes in performance. KR was frequently mentioned as a necessary condition for efficient learning (Embrey, 1979). Studies have documented the usefulness of KR in improving inspection performance, but timely and relevant KR is unusual in the actual inspection environment. In addition, other steps such as obtaining supervisor and trainee motivation, identifying training needs, developing training programs, and evaluating their effectiveness are also important in any learning strategy.

The basic task of the industrial inspector is straightforward: to search a prespecified area, compare each event with one's mental "defect" model, make a decision on its acceptability within established quality limits, and take some kind of action based on the decision. It is a much more complex issue, however, to be able to completely predict inspector performance from known parameters. An exact model of human inspection behavior has not yet been developed, but it is now possible to avoid accepting unrealistic assumptions of operator performance and to predict performance changes based on the effects of many inputs to the human inspector (Drury and Fox, 1975). Both search models and decision-making models, developed from human engineering data and theories, display a certain amount of predictive and operational utility. In addition, cases of prolonged periods of inspection also require vigilance models to predict performance.

Vigilance Behavior

The length of the inspection period is an important factor in predicting overall performance. Many studies have clearly shown that defect detection declines as a function of time (Mackworth, 1964). This so-called "vigilance decrement" represents a general deterioration in performance during extended monitoring tasks. This decline is quantifiable in terms of both a decrease in the number of signals detected and an increase in response latency. The deterioration can be rapid, with drops as much as 40% in 30 minutes reported (Fox, 1975). Although some researchers maintain that the vigilance decrement is a laboratory artifact (Smith and Lucaccini, 1969), sufficient industrial evidence exists to justify concern (Poulton, 1973). Fox (1975) recommended frequent rest breaks and job enlargement as two techniques to reduce the vigilance effect.

While the causes of the decrement are largely unknown, investigators have started to consider the cognitive demands of vigilance behavior. Williams (1986) suggested that inadequate training and taxing information processing demands are among the possible sources of the decrement. Inadequate training results from the failure of operators to adopt a stable response criterion for judging items as "defects" or "nondefects" prior to testing. Operators who were initially over responsive with signal reports gradually decreased their frequency of reported signals to correspond more to the actual frequency with which signals are presented. This probability matching strategy is the result of both training and feedback on the event sequence structure. On the other hand, high processing demands brought about by memory load and time pressure (Parasuraman, and Davies, 1976) also reduced observer sensitivity. Williams (1986) examined vigilance performance while compensating for the effects of both these proposed sources of error. The results indicated that the training scheme for stabilizing response bias by using a probability matching strategy prior to testing was successful; however, sensitivity continued to decline with event rates as low as 20/minute. This is evidence that the 24/minute event rate cutoff between high and low event rates (Parasuraman and Davies, 1976) needed revision. Although the role and extent of vigilance effects during visual inspection are not completely known, the scientific study of the factors will benefit training for both vigilance and inspector behavior.

Inspection Performance Measures

Before developing training techniques or learning strategies for visual inspection, sensitive and relevant performance measures are required. The three primary dependent measures used in vigilance research included correct detection rates, false alarm rates, and response latencies (Davies and Parasuraman, 1982). While false alarm rates were only reported sporadically, it was not until the application of decision theory that a satisfactory way of combining these two measures became available.

Correct Detections

Correct detection of defects is the most frequently used measure of inspector sensitivity. However, while detecting defects is the prime objective of inspectors in general, it confounds inspector sensitivity with decision bias. For example, an inspector who is completely ignorant of the differences between defects and nondefects can, nevertheless, achieve 100% defect detections by responding positively ("defect") on every trial (Davies and Parasuraman, 1982). The inspector, in this example, was biased to respond "defect" more often than "nondefect" on a given trial. Without a measure that also accounts for decision bias, inspector sensitivity is easily overestimated.

Likewise, studies which only used the number of missed signals or just false alarms to measure performance also suffered from the same inability to account for response biases of human inspectors. It wasn't until the development of Signal Detection Theory that both sensitivity and response bias of inspectors could be separately analyzed.

Signal Detection Theory (SDT)

Most visual inspection tasks consist of search followed by decision making. In order to study the effects of training on inspector decision making, it is necessary to either completely account for observer biases in a search model (Grindley and Townsend, 1970) or minimize, to the greatest possible extent, the requirement to search. When the latter condition has been met, it is possible to use SDT to model the decision-making component of visual inspection (Wallack and Adams, 1969; Drury and Addison, 1973; Drury, 1975). The SDT model was first applied to separate out the physical and psychological aspects of signal detection.

All human information processing begins with the detection of some stimulus event in the environment. This event may be a change in brightness of a light source, a change in frequency of an auditory signal, a tumor on an X-ray, a defect on a circuit board, or an enemy target on a radar scope. Signals are always detected against a background of noise which produces observer errors. Perfect detection performance is unusual and errors often involve more than just a lack of sensory acuity (Lachman, Lachman, and Butterfield, 1979).

During the years prior to the 1950s, many psychophysicists were busy measuring the detectability of signals as a function of intensity for various modalities. The classic threshold model was developed during this time for specifying those signal intensities at which the subject correctly detected a certain percentage of signals (usually 50%) (Van Cott and Kinkade, 1972). A psychometric function was generated by plotting signal intensity versus percent correct responses for various types of signal input. Interestingly, different functions could be obtained by simply giving different instructions to a single subject such as, a) "Be sure not to miss any signals," b) "Just detect as many signals as you can without worrying about it," or c) "Be absolutely sure a signal is present before responding" (Van Cott and Kinkade, 1972). These instructions can change a subject's threshold for a particular stimulus intensity. As a result, the conceptual meaning of "threshold" as purely a function of the physical properties of both the stimulus and the observer must be altered. A more subjective component must also be included to reflect, among other things, instructions given and the response bias of the observer. Therefore, a more sensitive model is needed to account for both these components of human signal detection.

The SDT model was developed to separate the relative effects of observer sensitivity and response bias on detection performance (Green, 1960; Swets, Tanner, and Birdsall, 1961). This model assumes that there are two stages of information processing during signal detection tasks: 1. Sensory information is accumulated concerning the presence of a signal, and 2. a decision is made whether this evidence constitutes a signal or noise (Wickens, 1984). By dividing the world into discrete states (signal and noise) and allowing the observer only two responses (yes or no), the set of all possible outcomes can be specified in a 2 X 2 matrix (see Figure 2).

		STIMULUS	
		NOISE (NONDEFECT)	SIGNAL + NOISE (DEFECT)
RESPONSE	"YES"	FALSE ALARM P(Y/N)	HIT P(Y/SN)
	"NO"	CORRECT ACCP P(N/N)	MISS P(N/SN)

Figure 2. SDT Decision Matrix

Visual inspection studies have often conceptualized the inspection task in terms of a human observer's ability to detect signals embedded in noise (Wallach and Adams, 1969). Though sometimes criticized for being too nonrepresentative of real industrial-based tasks, SDT has been used to analyze inspector performance both in the laboratory (Embrey, 1975) and in the field (Drury and Addison, 1973). SDT considers the quality control inspector to be a statistical hypothesis tester, gathering data from each observation and deciding if a particular item was sampled from a distribution of defects or a distribution of nondefects. Due to continuous variation in noise underlying both these distributions, some defects will be missed and some nondefects will be judged to be defects. The inspector's sensitivity, or d', is defined by a joint consideration of missed defects and falsely judged defects (Green and Swets, 1966):

$$d' = (u_d - u_n) / \sigma \qquad (1)$$

where: d' = Inspection Sensitivity
u_d = Mean of Defect Intensity Distribution
u_n = Mean of Nondefect Intensity Distribution
σ = Standard Deviation of Intensity Distribution

This measurement theory assumes that the variances of the two intensity distributions are identical, and that the evidence distribution of defects and nondefects are both normally distributed. When these assumptions are not met, additional adjustments to the data may be necessary to avoid confounding sensitivity and response bias. Several nonparametric measures were discussed by Green and Swets (1966). While the concept of an evidence distribution is somewhat abstract, this is usually meant as random variations in product quality along the inspected dimension, as well as noise within the inspector. For example, if one is judging whether a portion of a circuit board has been scratched, there will be a distribution of small scratches (in width, length, and depth) on a non-defective board. The defective board will also contain a distribution of scratches, but the mean width, length, and depth of scratches will be greater. The second performance parameter is ß, or the inspector's response criterion (bias) in making a decision. ß is one of several ways to represent the relative position of one's criterion along the evidence dimension. It is calculated as likelihood ratio of the defect over nondefect probabilities for a particular criterion:

(2) $$\beta = y_d/y_n \text{ where:}$$

ß = Inspector Response Criterion

y_d = Ordinate of Defect Intensity Distribution at Inspection Decision Criterion

y_n = Ordinate of Nondefect Intensity Distribution at Inspection Decision Criterion

If an inspector behaves in accordance with SDT, d', which is based on the effective signal strength or discriminability of the defective items, should remain constant over the inspection period. ß, on the other hand, should vary depending on the defect probability and perceived values and costs of decisions at particular points in time (Drury and Addison, 1973). ß and d' are assumed to be independent measures of visual inspection performance. Figure 3 illustrates the theoretical probability density functions for the evidence variable of events classified as either signals (defects) or noise (nondefects).

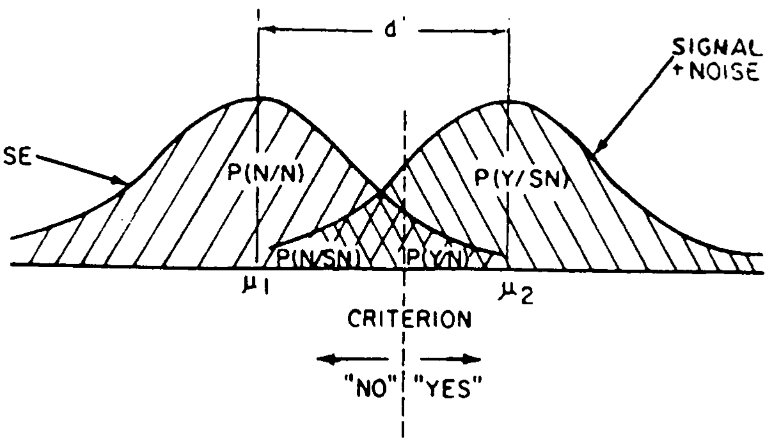


Figure 3. Hypothetical SDT Distributions. Source: Van Cott and Kinkade, (1972)

SDT, as a normative model, can also prescribe the optimal value of ß, known as ß*, which maximizes decision values while minimizing error costs (Green and Swets, 1966). Both defect probability and the values and costs of decision outcomes must be represented in a model of β^* performance. β^* is computed by (Swets, Tanner, and Birdsall, 1961):

(3) $\beta^* = [P(ND) * (VCA + CFA)]/[P(D) * (CMISS + VHIT)]$

where: P(ND) = probability of a nondefect
VCA = value of a correct acceptance of a nondefect
CFA = cost of a false alarm
P(D) = probability of a defect
CMISS = cost of a miss
VHIT = value of a hit

If the values and costs of decision outcomes are the same (symmetrical payoff matrix), the equation for β^* is reduced to:

(4) $\beta^* = P(ND)/P(D)$

where the two probabilities are those that actually exist during the inspection task.

An inspector may adopt a liberal criterion (small β) which maximizes both correct and false detections, or he may adopt a more conservative criterion (large β) which minimizes both of these. The degree of inspector optimality can be assessed by computing the absolute value of the difference between an inspector's actual β and β^* for a given defect probability and payoff matrix:

(5) Degree of inspector optimality = $|\beta - \beta^*|$

Studies have shown that observers are generally "sluggish" in shifting their ß's as a result of changing probabilities and payoffs (Green and Swets, 1966). For example, observers tend to be less liberal then they should be for small β^* values, and less conservative then they should be for large β^* values (Peterson and Beach, 1967). This inherent "conservatism" may be the result of an overall inability of human observers to accurately combine the diagnostic meaning of several pieces of data when revising probabilities (Edwards, 1982).

In real world applications, operators do shift their ß's in the required direction in response to changing probabilities, although not as far as dictated by the β^* model. Drury and Addison (1973) found that quality control inspectors examining sheet metal for defects will adjust their ß's according to the estimated defect rate of the batch. In addition, Wickens (1984) reported the results of a study which applied the SDT model to the air traffic controller's task of deciding whether the merging paths of two aircraft signal a collision (Bisseret, 1981).

Controllers lowered their ß (became more willing to specify a correction) as the difficulty of the task increased. Furthermore, experts were more likely to set their ß at a lower value than trainees. The author suggested that trainees are more uncertain about how to implement a correction and, therefore, more reluctant to call for a correction. Thus, response criterion training could improve performance of trainees in the air traffic control environment.

Wallack and Adams (1969) pretrained industrial inspectors to detect nicks in stranded electrical conductors. Four levels of product percent defective were used with d' and ß values calculated for each. The greatest difference between ß and ß* occurred at the lowest percent defective level (5%). In addition, two distinct populations of inspectors in the 5% defective group could be distinguished on the basis of Type 1 and Type 2 errors. Although there are problems associated with this type of joint laboratory-industrial environment research (Adams, 1975), the benefits of validating these models in the actual inspection workplace outweigh any methodological cost.

Response Latencies

The amount of time it takes an inspector to make a decision is also an important measure of performance. Buck (1966) reported that there was evidence of an inverse relationship between detection rate and latency; observers who detected more defects also made faster decisions. This result dismissed an alternative explanation that greater defect detection is due to longer observation times (speed-accuracy tradeoff). Thus, if some central process called "vigilance" mediates performance, then detection rate and latency may be related if both reflect changes in vigilance (Davies and Tune, 1970). In addition, if higher response times also reflect increased levels of uncertainty, then inspector latency represented changes in response criterion as well as changes in sensitivity.

Training for Visual Inspection

Traditionally, training research has focused primarily on motor learning rather than the perceptual skills required in product inspection tasks (Welford, 1968). Perceptual learning involves covert mental processes which are sometimes inaccessible to the trainee. As a result, it is difficult to operationalize these cognitive variables in an experimental setting to measure the effects of training. Without these data, it is difficult to establish a link between theoretical models and the actual inspection environment.

Training Effects on SDT Parameters

Plagued by many of the same problems encountered by those studying human monitoring behavior, visual inspection tasks have been analyzed using SDT where the effects of observer sensitivity and decision bias can be separated and analyzed (Baker, 1975). While the SDT model is a useful tool for understanding inspection performance, there is still scant research on the training effects in the model parameters. Most studies which have investigated the effects of training on inspection performance have focused on enhancing the inspector's sensitivity through KR or cueing techniques (Embrey, 1979). Equally important, however, is the impact of training on the response criterion set by the inspector during the actual task. As measured by β, the response bias provides evidence for the accuracy and completeness of an inspector's internal model of the process. Any research attempt to model inspector decision making must integrate these two characteristics of inspection performance.

The first, observer sensitivity, is a function of such factors as visual acuity, discriminability of defects and nondefects, and the observer's knowledge of defect characteristics. The inspector's response criterion, on the other hand, reflects an individual's rule for making inspection decisions based on the a priori defect probability and the values of costs of various decision outcomes. In addition, inspector performance can also be assessed in terms of an individual's reaction to changes in the defect probabilities occurring during the task itself. The innate conservatism of an observer, together with his/her limited sensitivity, can produce subjective estimates of defect probabilities which lag behind actual probabilities. Based on these observations, Embrey (1975) recommended three training objectives for visual inspection:

1. Inspector sensitivity should be maximized for a given defect.

2. The response criterion adopted by the inspector should be compatible with the ongoing defect probability and the costs and values associated with the decisions.

3. The inspector should be able to modify his/her

response criterion in accordance with changes in defect incidence or the costs and values of decisions.

Training and Skill Retention

The best way to judge the effectiveness of training is to measure the amount of time with which a certain level of skill is maintained on a task. The retention interval, here, refers to the period of time after training during which subjects do not perform the task. This can occur, for instance, when an inspector is rotated through several stations where substantially different defects must be detected. Several factors have been identified which influence the retention of a particular skill.

Training Duration

The effects of training duration were explored in several studies cited by Hagman and Rose (1983). Subjects who performed more repetitions of a 52 step procedure involving testing of alternator electrical output had faster performance times and less errors immediately after training and two weeks later. In the second study, subjects who were trained to a 'mastery' criterion level for assembly/ disassembly of an M-60 machine gun required fewer trials and made fewer errors to relearn the task to proficiency (error free performance) after eight weeks than a comparable group of subjects exposed to half as many trials. In the third study, a "mastery" subject group trained to a criterion of three consecutive error-free trials of boresighting and zeroing the main gun of the M60A1 tank retained the procedural skills better (based on the number of errors committed on the first trial after retention) than a comparable group trained to a criterion of one error-free trial.

Naylor, Briggs, and Reed (1968) also varied the duration of training for subjects learning to perform a three dimensional tracking task and a procedural secondary task. Subjects who spent more time in training had fewer errors and performed better at both levels of the secondary task after both one and four weeks of retention.

Distribution of Training

Another factor that affects training is the timing of additional trials. Hagman and Rose (1983) reported the results of a study where two groups of reservists were trained in assembly/disassembly of machine guns. One group received extra repetitions during initial training while the second group received their additional trials after four weeks. Both groups committed the same number of errors and required the same number of trials to attain proficiency when tested eight weeks after initial training.

Massed versus spaced repetitions of training trials is another way of varying distribution of training. Hagman and Rose (1983) reported the results of a study where two groups of subjects were trained on a task of testing alternator output using three massed or three spaced repetitions prior to retention testing two weeks later. The massed training group took longer and committed more errors than the spaced training group. The advantage of spaced repetitions of training is a fairly consistent result throughout the training literature.

The Retention Interval

Roehrig (1964) reported near perfect retention for subjects who were able to perform a simple balancing task at pre-retention performance levels after not practicing for 50 weeks. Performance continued to improve with additional trials as though there had been no retention interval at all.

Fleishman and Parker (1962) trained two groups of subjects on a complex compensatory tracking task, one group was trained by rote practice without feedback while the second group received instructions and feedback on performance. Both groups were tested at various retention intervals ranging from 1 to 24 months. Only feedback trained subjects showed no performance decrement for any retention interval. The authors concluded that the important fact in retention was not the type of training administered but the level of proficiency attained. Naylor, Briggs, and Reed (1968), on the other hand, reported significant reductions in tracking performance for a four week retention interval compared to one week.

Retention factors in visual inspection have not been adequately explored. Most research available addressed retention during motor rather than perceptual learning. Inspector training must be evaluated from the standpoint of retention of skill as well as performance measures after training. Despite the importance of retention factors (especially long-term retention) to both motor and perceptual skill learning, little work is being done and few new ideas generated (Adams, 1987).

Task Difficulty and Inspector Training

Overall task performance can usually be improved by first training workers on smaller components of the task (Wightman and Lintern, 1985; Schneider, 1985). Therefore, deciding which components should be trained and in what order become important considerations. One method for selecting the task components to be trained is based on relative workload effects. According to a resource view of attention (Moray, 1967; Kahneman, 1973), skilled task performance requires the investment of limited processing resources which must be allocated in greater amounts as the demands of the task increase. Within this context, workload is related to the amount of processing resources demanded by a task compared to those supplied by the operator. The limited availability of these resources combined with their multiplicity (Wickens, 1980) can have serious implications for training complex skills. As the difficulty of a task increases or as concurrent tasks compete for the same processing resources, the higher the task workload and the more resources needed to maintain performance. On the other hand, increasing the difficulty of some tasks does not increase workload and the further investment of resources benefits neither performance nor learning. The first type of task, known as "resource limited," forces the trainee to invest more resources as task difficulty increases, improving performance. The second type, "data limited," describes tasks whose performance remains unchanged despite increasing task difficulty. Mane and Wickens (1986) hypothesized that increased levels of task difficulty will facilitate learning when these increases are both resource loading and derived directly from task learning. Conversely, when workload stems from the need to perform another task or aspects of a task that do not benefit learning, the learning of that task component will suffer.

At first glance, this idea that training on high difficulty tasks will improve post-training performance may seem at odds with the principle of adaptive training, where task difficulty is varied as a function of how well the trainee is doing (Kelley, 1969). Under this approach, a trainee would start out with a relatively easy version of the task to be trained and then transferred to a more difficult version once performance met some criterion level. The assumption here is that there should be a positive transfer from easy to difficult tasks. Mane and Wickens (1986) stated that such positive transfer would occur when the task is data limited. On such a task, increasing the workload, and therefore the processing resources involved, does not affect performance. On the other hand, resource limited tasks, where performance improves as the amount of processing resources increases, would experience positive transfer from difficult to easy versions of a task.

Therefore, the level of task difficulty is an important factor to consider when designing training programs for visual inspectors.

Training Inspectors' Internal Models

Many complex human behaviors are thought to be guided by an individual's "internal" representation of the environment. This representation can be described in terms of an internal or mental model of some physical process or activity which operators can use as a basis for understanding and predicting the response of a human-machine system (Wickens and Kessel, 1979). Many have accounted for important human performance changes in terms of certain selected parameters of an operator's internal model. Veldhuyzen and Stassen (1976) observed that all forms of human behavior require some internal representation of the system being observed or controlled. For example, human monitors continually compared their internal model to the actual system until the observed difference exceeded some subjective criterion and a "failure" is detected (Wickens and Kessel, 1979).

Veldhuyzen and Stassen (1976) acknowledged that predictions based on an internal model were not always accurate since:

1. The structure of the internal model may differ from the structure of the system to be controlled or monitored.

2. The internal model parameters may differ from the parameters of the system to be monitored or controlled.

3. The system can only be perceived with restricted accuracy.

4. Disturbances are often not known exactly.

Bayesian decision theory has been used to formalize and externalize a decision maker's internal model; however, Tversky and Kahneman (1974) cautioned that people use nonoptimal, stereotypical models of probabilistic processes in estimating the likelihood of events. These inherent inaccuracies and limitations of a human operator's internal model may be used to establish important model parameters needed in a specific training environment.

Although few have measured and manipulated internal model parameters during training, numerous investigations focused on the more generalized topic of training and decision making performance. Wickens (1984) described three types of decision-making aids that have been shown to be useful. First, make the decision maker aware of unconscious biases that may be influencing performance. For example, Rouse and Hunt (1981) succeeded in improving diagnostic performance by training subjects to extract information from the absence of failure information. Second, provide accurate and timely feedback to decision makers so that they are forced to judge and evaluate the success or failure of their rules. Tversky and Kahneman (1973) recommended that decision makers should be trained to encode events as probabilities rather than frequencies, since probabilities inherently account for both positive and negative evidence. Finally, the correlational structure existing in the cues that represent a certain hypothesis should be emphasized. Humans have shown a consistent ability to integrate cues when correlations are known ahead of time.

Recent work on internal models has been concerned with extremely complex physical systems or with behavior in ill-defined tasks such as how an electrical circuit works (Gentner and Stevens, 1983). The internal model is also a hypothetical construct which can account for several aspects of process control behavior. First, an internal model is thought to guide the display sampling and scanning of multifunction systems (Moray, 1981). It can also formulate plans of action and translate intended goals into present control actions. Finally, the internal model forms the source of the operator's expectancies of the relationships between variables.

Kieras and Bovair (1984) investigated the role of internal models in learning to operate a relatively simple device. Their objective was to demonstrate that providing a device model during training can result in faster learning and better retention of operating procedures. The results showed that the device model trainees learned the procedures sooner, executed them faster and retained them more accurately than the no-model trainees. Device model trainees were also more able to infer operating procedures. This advantage was due to the specific configuration of components and controls present in the model and not to the motivational aspects, component descriptions or general descriptions provided by the model. These results supported their recommendations concerning when and what kind of device model information should be taught to operators; however, no details on the structure of the operator's internal device model were provided.

Although the concept of a internal model may seem straightforward for learning to operate an external device, it is more difficult to apply to a cognitively complex task such as visual inspection. In detecting a visual target, it is hypothesized that an inspector uses some kind of internal model of the inspection environment to make decisions. For example, one way to conceptualize this model is in terms of an SDT framework where both specification (as measured by d') and probabilistic (as measured by ß) information would be important components. As already discussed, the SDT model is a useful tool, not only for describing inspector performance in terms of d' and ß, but also for prescribing normative (optimal) behavior for a given set of external factors. Training for inspection, therefore, can be viewed as either "providing" a valid internal model to trainees or "optimizing" the existing models of current industrial inspectors. The discrepancy between actual and optimal ß can monitor the progress of internal model development during training or to assess the quality of an inspector's internal model at the end of training. This type of analysis is also useful for understanding the distinction between novice and expert performance. The evolution of knowledge from novice to expert levels begins during training and is logically related to the development of an inspector's internal model.

Knowledge of Results (KR)

As a common technique for promoting perceptual learning (Annett, 1966), KR is knowledge received relating to the outcome of one's responses (Wiener, 1968). Commonly used during vigilance tasks (Antonelli and Karas, 1967; Warm, Epps, and Ferguson, 1974), KR has also been used as an effective technique for providing defect specification and distribution information during visual inspection (Drury and Addison, 1973; Embrey, 1975). One's increased sensitivity resulting from KR has been ascribed to either increased motivation or enhancement of defect knowledge (Embrey, 1979; Mackworth, 1964). In addition, higher sensitivity may also allow more optimal adjustment of ß by providing more correct opportunities from which to estimate the true defect rate (Williges, 1973). Thus, KR may allow the development of a more optimal response strategy as defect probabilities change. To better address the scope of KR effects, Embrey (1975) presented KR and several combinations of signal probabilities to inspectors detecting changes in brightness of a central disk. Although KR increased subjects' d', independent of signal probability, subjects' ß's were more optimal in the NO-KR condition as KR lowered ß more than predicted. While the author did not try to interpret this result in terms of cognitive theory, it is apparent that KR had inconsistent effects on inspection performance. In order to reconcile such results, a cognitive model of KR utilization will be developed to describe and explain KR's influence on decision making during inspection.

ß Variability

One possible explanation for KR's influence on sensitivity is in terms of changes in the variability of an inspector's response criterion (Drury, 1988). For example, Figure 4 shows a Receiver Operating Characteristic (ROC) curve (Green and Swets, 1966) with two different criteria, $ß_1$ and $ß_2$. If an inspector divides his time equally between each criterion, then the expected value of his criterion will be at some point along the line joining $ß_1$ and $ß_2$. Any point along this line represents a lower sensitivity than either $ß_1$ or $ß_2$. Therefore, the lower the criterion variability ($ß_1$ or $ß_2$), the higher the apparent sensitivity. If it is assumed that KR provides an inspector with the knowledge of his own response criterion, then it's possible he uses it to reduce variability. As a result, KR's increase in sensitivity would become an artifact and not the result of information transfer.

To test this hypothesis, the hit rate and false alarm rate for each block were plotted, giving three points on the ROC curve (see Figure 4). The smallest sum of the straight line distances connecting these three points on the ROC curve was calculated to represent ßVAR. In Figure 4, ßVAR is represented by the straight line distance connecting $ß_1$ and $ß_3$ plus $ß_3$ and $ß_2$. ROC points spread out along the curve, representing high ßVAR, also had large intra-point

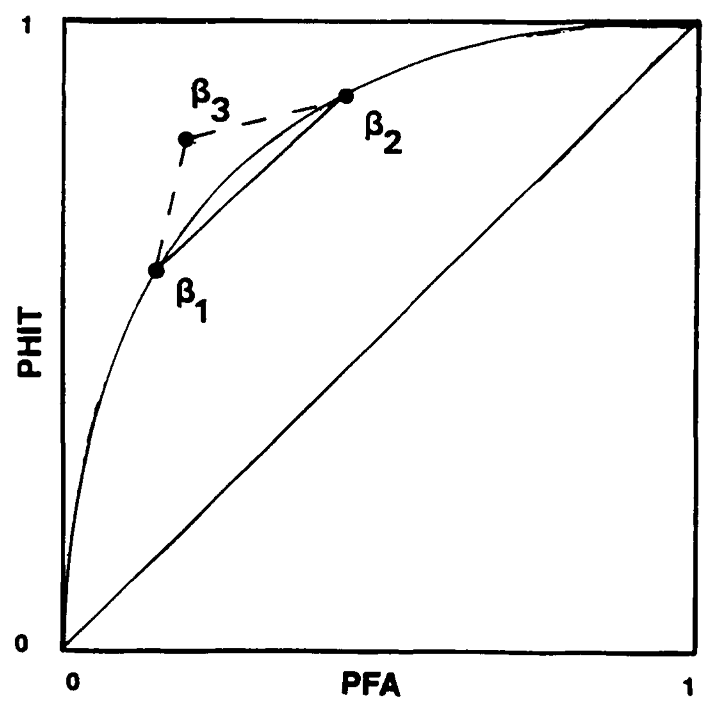


Figure 4. Theoretical ROC for ß Variability

distances; those points packed more closely had lower ßVAR and smaller intra-point distances. This measure was used in Experiment 1 to test whether inspector sensitivity was related to ß Variability.

A Model for KR Utilization

Most decision-making models for visual inspection have been based on normative theory borrowed from the mathematical and physical sciences (Drury, 1975). Many of these models included feedback to the decision maker. However, the underlying cognitive processes that manipulate and utilize this feedback are usually not represented. An understanding of these processes is vital to accurately model the decision making behavior of visual inspectors. Figure 5 illustrates a cognitive model for KR utilization during inspection based on the research of Sternberg (1967,1969), Wallack and Adams (1969), and Adams (1975), among others. In this model, inspector knowledge, consisting of defect characteristics and probabilities, as well as the perceived values and costs of decisions, is assumed to be represented in a veridical format within memory which preserves spatial information (Embrey, 1979). The ability of the inspector to estimate the respective

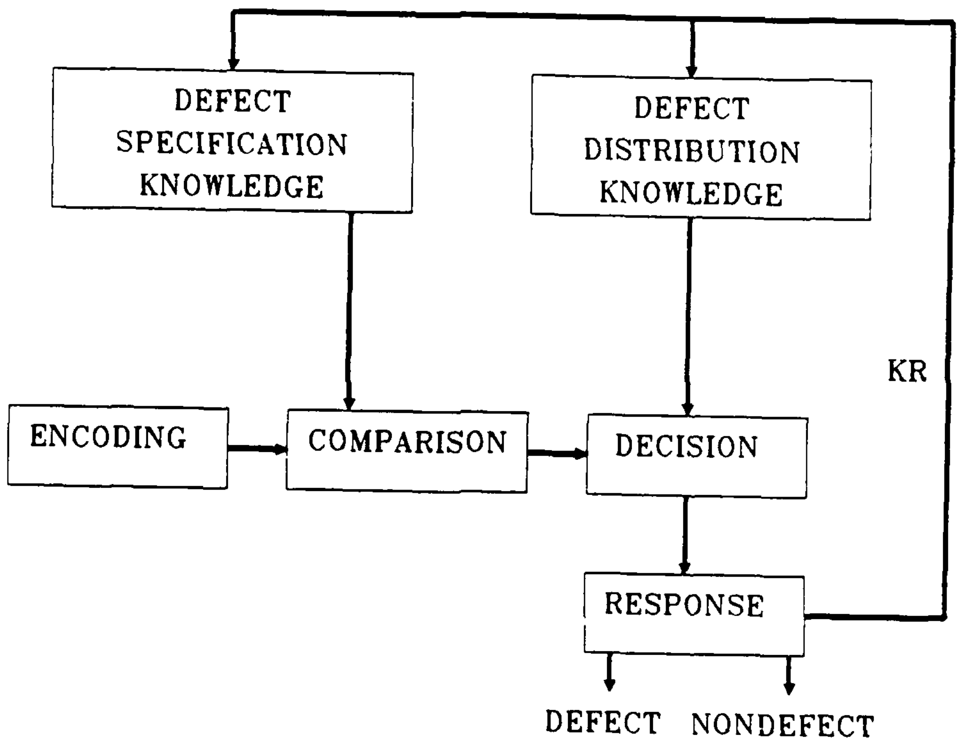


Figure 5. Initial KR Utilization Model

means of both the defect and nondefect distributions is represented in the model as inspection specification knowledge. On the other hand, the ability to estimate defect and nondefect probabilities and the actual costs and values of decisions is represented as inspection distribution knowledge. Based on this interpretation, d' reflects the level of specification knowledge and ß reflects the level of distribution knowledge.

KR can influence inspector knowledge directly by confirming or disconfirming hypotheses about the characteristics and distribution of defects versus nondefects. Defect characteristics are learned better with KR since inspectors are now aware of errors on specific trials. Each error forces the inspector to update his/her defect model resulting in higher performance compared to no KR with fewer updates. KR also provides evidence of the event sequence structure during an inspection period which can be used to estimate defect probabilities and adjust ß. Overall inspector performance, therefore, should be enhanced via KR both in terms of sensitivity and response bias, and should be independent of the relative difficulty of the inspection task. In addition, during sudden shifts in the defect prevalence (e.g. a process breakdown causing an increase in defects), inspectors should adjust ß more optimally with KR due to superior defect distribution knowledge. These predictions will be tested in the following series of experiments.

Chapter 2

OBJECTIVES AND HYPOTHESES

Objectives

The overall goal of this research is to develop a model of KR utilization for decision making during visual inspection. To obtain sufficient evidence to support such a model, three experiments with the following objectives were conducted.

Experiment 1

1. The effects of KR on inspector sensitivity, response bias, and optimal ß placement both within and between defect probability conditions.

2. The effects of task difficulty, sequence of difficulty levels, and defect probability on inspector performance.

3. The relationship between d' and other dependent measures, including ß variability, both within and between inspection groups.

Experiment 2

1. The relative effects of True versus False KR on both inspector sensitivity and response bias across specified payoff and probability conditions.

2. The effects of changing values and costs of decision outcomes on inspection performance.

3. The effects of increasing defect probability on inspection performance.

Experiment 3

1. The effects of training with KR on inspection sensitivity and response bias when KR is no longer available.

2. The effects of task difficulty during training on subsequent inspection performance.

3. The effects of increasing defect probability during training and subsequent phases on inspection performance.

4. The effects of inspector training on post-test performance both immediately after training and 3 weeks later.

Hypotheses

Based on the previous objectives and the initial model already developed, hypotheses were generated which predicted the effects of the manipulations on the dependent variables used in the three experiments.

Experiment 1

Inspectors given KR should have:

1. higher sensitivity as measured by d'.
2. more optimal response criteria as measured by $|\beta - \beta^*|$.
3. faster RT's.

compared with NO-KR inspectors.

Decreasing discriminability of defects (higher difficulty) should result in:

4. lower d'.
5. lower β.
6. slower RT's.

Increasing defect probability should:

7. have no effect on d'.

8. decrease ß.

9. have no effect on RT.

Overall inspector d' should be:

10. unrelated to ß.

11. unrelated to |ß-ß*|.

Inspectors performing the Low to High Defect Discriminability Sequence should have:

12. higher d's.

13. larger |ß-ß*|'s (less optimal).

14. faster RT's.

compared with inspectors in the High to Low sequence.

Experiment 2

Inspectors provided with TRUE-KR should have:

1. higher d'.

2. lower |ß-ß*|.

3. and faster RT's.

compared with either FALSE-KR or NO-KR inspectors.

Inspectors provided with FALSE-KR should have:

4. higher d'.

5. lower |ß-ß*|.

6. faster RT's.

than NO-KR inspectors.

Manipulating ß by changing decision payoffs to reinforce correct defect detections should result in:

7. constant d' .

8. lower $|ß-ß^*|$.

9. constant RT.

compared to changing defect probabilities.

Experiment 3

Inspectors trained with KR should have:

1. higher d'.

2. lower $|ß-ß^*|$.

3. faster RT's.

than their NO-KR counterparts during training, immediate retention, and three week retention intervals.

Inspectors trained with KR and High Difficulty defects should have:

4. higher d'.

5. higher $|ß-ß^*|$.

6. faster RT's.

during immediate and three-week retention intervals.

These predicted effects are summarized in Table 1.

Table 1. Summary of Predicted Effects of Experimental Manipulations

EXP	Contrast Difference	Dependent Variable d'	ß	|ß-ß*|	RT
1	KR - NO-KR	+*	ND	-	-
	High - Low Difficulty	-	-	+	+
	0.2 - 0.4 Probability	ND	+	ND	ND
	Low to High - High to Low Difficulty Sequence	-	ND	ND	ND
2	TRUE KR - FALSE/NO-KR	+	ND	-	-
	Symmetric - 2X/9X Payoffs	ND	+	ND	ND
	0.2 - 0.4 Probability	ND	+	ND	ND
3	Phase 1				
	KR - NO-KR	+	ND	-	-
	High - Low Difficulty	-	-	+	+
	Phase 1 - Phase 2				
	KR - NO-KR	+	ND	-	-
	High - Low Difficulty	-	-	+	+
	Phase 2 ---> Phase 3 (3 weeks)	(No Significant changes)			

* + = positive difference between the two manipulations
- = negative difference between the two manipulations
ND = no difference

Chapter 3

EXPERIMENT 1: KNOWLEDGE OF RESULTS IN VISUAL INSPECTION DECISIONS: SENSITIVITY OR CRITERION EFFECT?

The overall objective of this experiment was to assess the effects of both KR and task difficulty on the SDT parameters as defect probability increased during visual inspection.

Method

This section will highlight the major equipment and personnel requirements for performing this experiment. The experimental design is discussed with a detailed explanation of the procedure.

Subjects

Twenty right-handed males, recruited from a local newspaper ad, participated in this experiment. Each subject was paid up to $5.00 per hour, including performance incentive pay, for 1.5 hours of experimental time. All subjects were screened for 20/20 or better corrected visual acuity and ranged in age from 16 to 41 years.

Apparatus

The visual inspection task was presented on an AT&T PC-6300 personal computer located in the Industrial Engineering/Human Factors Laboratory at The Pennsylvania State University. The monitor used in the experiment was an 11.5" AT&T color monitor (model CRT318H) with EGA capability, located 20 inches from the subject's eyes. A mouse (Logitech Model No. P7-2F-AT) was used to record all subject responses, using the two buttons on top of the mouse as response keys. The intensity of the monitor was 16 foot-lamberts with a contrast of 86%. Ambient illumination at the task was 30 foot-candles.

Visual Inspection Task

An inspection task was created on the screen of the monitor by mimicking well known work standards for circuit board metallization scratches (Martin Marietta, 1981). According to the work standards, a scratch across an etched conductor on a circuit board is acceptable if more than half the width of the conductor is left undisturbed (Martin Marietta, 1981, p. 1-3). The required perceptual skill here is a visual discrimination of distance, comparing the relative width of the scratch with the width of a conductor.

Figure 6 displays the analogous experimental task, which required subjects to judge the length of line segments (i.e., "scratches") displayed on the screen. A line was defective if it extended more than half way across the width

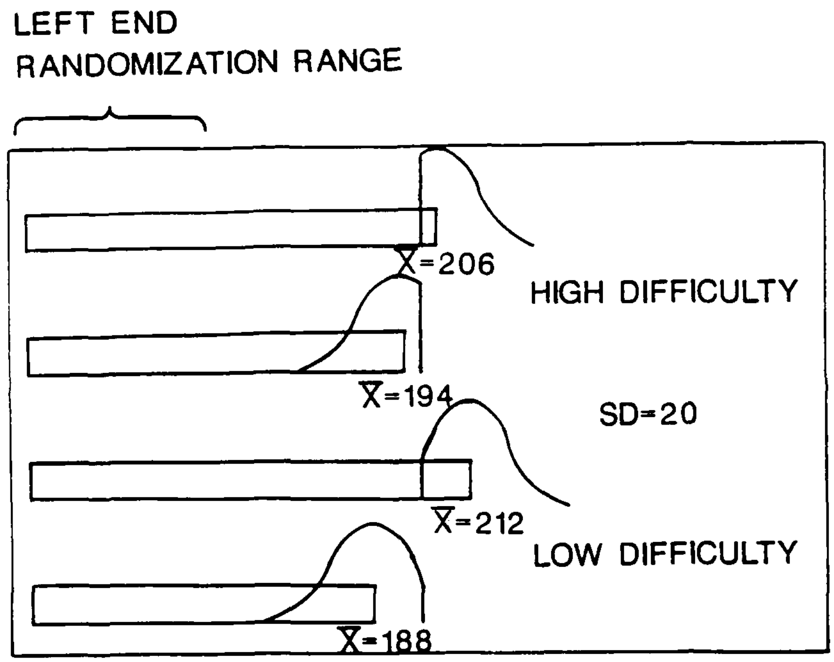


Figure 6. Visual Inspection Task

of the viewing area (i.e., "conductor"), and was nondefective otherwise. The line segments were randomly presented along the imaginary centerline connecting the two vertical sides of the viewing area. Since the line segments were always presented along this centerline, the inspection task did not require visual search but only addressed inspector decisions. The relative horizontal position of each line segment was varied from trial to trial by randomly selecting the starting point for drawing the leftmost end of the line within the range of 25% of the viewing area width.

Each line segment was presented on the computer screen for 2 seconds. Subjects could respond during this 2-second exposure time or any time during the five second interval immediately following the removal of the line segment. Subjects responded by pressing either the right-hand button on the mouse for a defective line segment or the left-hand button otherwise. If seven seconds passed without a response, a miss was recorded, followed by the next line. There was always a 2-second pause between the subject's response and the presentation of the next stimulus. The computer recorded responses and decision times for subsequent analysis.

Experimental Design

Four fixed independent variables, with subjects randomized, were utilized in this experiment, as schematized in Table 2 and described below. Although generalizing beyond the specific levels included here will be difficult,

Table 2. Experiment 1 Design

Discriminability Sequence	Defect Probability	NO-KR	KR
		Discriminability	
		LO --> HI	LO --> HI
		BLOCK I	BLOCK II
LOW to HIGH	0.2	150 trials	150 trials
	0.4	150 trials	150 trials
		Ss=1-5	Ss= 6-10
		BLOCK III	BLOCK IV
HIGH to LOW	0.2	150 trials	150 trials
	0.4	150 trials	150 trials
		Ss = 11-16	Ss = 16-20

NOTE: All subjects within a block performed both LOW and HIGH discriminability and 0.20 and 0.40 probability conditions.

the main intent is to study trends in performance from high to low difficulty (and vice versa), from low to high probability, and with KR or NO-KR using meaningful and relevant variable levels.

Knowledge of Results

Half of the subjects received KR and half did not. KR consisted of "right" or "wrong" statements at the top of the screen following each response. Summary accuracy and monetary reward information were also provided after each 50-trial block in the KR condition. Instructions to subjects in the KR condition also specified that the number of defects could change between blocks. No such information was given to those in the NO-KR condition.

Defect Discriminability

Manipulated within subjects, High Discriminability defects had an average 6% length difference between the two stimulus lines, based on 92% correct discrimination in pilot testing. Low discriminability defects had a 3% length difference, based on 75% correct discrimination in pilot testing. Both of these discrimination tasks were greater than a subject's expected threshold difference for line length judgement (Ono, 1967).

Discriminability Level Sequence

The order of discriminability levels described above was counterbalanced to avoid any carry-over or learning effects. Half the subjects started in the High defect discriminability task then moved to Low discriminability task, while the other half moved from Low to High Discriminability.

Defect Probability

The number of defects presented during each 50 trial block was always 10 (0.2 defect probability) in the first block and 20 (0.4 defect probability) in the second block of each discriminability condition.

Table 2 shows that KR and Discriminability Sequence were between-subject experimental manipulations (nested variables), whereas defect discriminability and probability were within-subject manipulations. Five subjects were randomly assigned to each one of the four between-subject conditions defined by KR and discriminability level sequence (see Appendix A for complete description of statistical model).

Procedure

A subject initially observed 20 trials with the half-way criterion line initially visible and held constant. Next, practice was provided, using two blocks of 20 trials (0.2 and 0.4 defect probability, respectively). Finally, each subject repeated the above practice session with performance recorded and evaluated to insure a performance level of at least 90% and 70% correct for High and Low Discriminability conditions, respectively. After training was successfully completed, a subject was administered three 50-trial block replications at each of the two defect probabilities. This was repeated for both Low and High Defect Discriminabilities for a total of 12, 50-trial blocks within each subject. Hit rate, false alarm rate, and mean reaction time (RT) were recorded for each block. Values for d' and ß were derived from the above data for subsequent SDT analysis. In addition, ß* and |ß-ß*| were calculated for each inspection condition.

ßVAR quantified ß variability using the three replication blocks as three points on the ROC. ßVAR was computed for each experimental condition and included in the subsequent ANOVA and regression analyses.

A bonus system insured a high level of subject motivation during the inspection task. Subjects were instructed that they could earn up to $3.60 in additional pay for performing at a consistently high level. For the KR condition, the bonus score was calculated on a block by block basis as Bonus = [P(hit)-P(false alarm)]*$.30. In the NO-KR condition, the bonus score was calculated after every third block as Bonus = [P(hit)-P(false alarm)]*$.90, but the amount was not revealed to the subject until the end of the experiment. A two-minute rest period was administered between each block of trials. Each subjects was paid at the end of his session.

Results

To obtain sufficient numbers of missed and false defects, the three blocks of 50 trials (150 trials total) were pooled within each discriminability condition for d' and ß computations. SDT analysis requires sufficient numbers of errors for accurate analyses, and pooling these data ensured this, at the expense of loss of resolution.

Hit Rate (HR) and False Alarm Rate (FAR)

Before deriving the parameters for the SDT model, an analysis was performed on HR, or probability of correctly calling a defect, and FAR, the probability of falsely naming a non-defect. Reported statistics below are from four-factor ANOVA's on each of the dependent variables. Duncan's Multiple Range Test (Montgomery, 1984) compared pairs of treatment means for significant main effects. Appendix B shows that the assumptions of the ANOVA model were met.

Hit Rate

The main effects of KR ($F[1,16]=5.12$, $p<.05$), Defect Discriminability ($F[1,48]=150.78$, $p<.0001$) and Defect Probability ($F[\_,48]=7.31$, $p<.01$) were all significant. HR increased with KR and decreased for Low Discriminability defects. In addition, HR was also higher in the 0.2 defect probability condition. There was also a significant KR X Defect Probability interaction ($F[1,48]=13.53$, $p<.01$), suggesting that KR provided a greater improvement in HR for 0.4 defect probability than the 0.2 probability.

False Alarm Rate

The only significant effect for FAR was the main effect of Defect Discriminability ($F[1,48]=92.63$, $p < .0001$). Not surprisingly, FAR was higher in the more difficult, Low Discriminability condition.

Sensitivity (d')

Table 3 contains mean values for d', ß, ß*, |ß-ß*|, and ß variability within each condition. Table 4 contains differences between condition means. Inspector sensitivity was an average of 0.47 greater in the KR condition when compared with NO-KR ($F[1,16]=7.22$, $p < .05$), as shown in Table 4. The d' from the High Discriminability task was an average 1.65 larger than from the Low Discriminability length difference judgement task ($F[1,48]=54.60$, $p < .001$), confirming the importance of task manipulations on d'. Though less significant, the sequence of defect discriminability levels was also important in influencing d'. Performing the Low followed by the High Discriminability task produced an average 0.38 greater d' than the opposite order $F[1,16]=4.65$, $p < .05$). This implied more efficient learning by starting on the more difficult task rather than the easier task. Interestingly, the Defect

Table 3. Mean Values for d', ß, |ß-ß*|, and ßVAR in Experiment 1

		Defect Discriminability							
		High				Low			
Defect	Prob	d'	ß	|ß-ß*|	ßVAR	d'	ß	|ß-ß*|	ßVAR
A	.2	3.27	5.67	3.89	0.20	1.48	2.00	1.87	0.22
	.4	3.32	3.52	2.26	0.11	1.35	1.79	0.63	0.27
B	.2	3.73	2.08	1.92	0.24	1.86	1.77	2.23	0.15
	.4	3.94	1.70	1.15	0.30	2.00	1.86	0.57	0.10
C	.2	2.90	6.93	5.64	0.25	1.52	2.49	1.51	0.21
	.4	2.52	7.50	6.17	0.18	1.21	2.19	0.71	0.32
D	.2	3.17	1.80	2.68	0.25	1.65	2.09	1.91	0.20
	.4	2.83	8.74	7.24	0.25	1.39	5.00	3.70	0.16

NOTE: A. KR, High to Low Discriminability Sequence
B. KR, Low to High Discriminability Sequence
C. NO-KR, High to Low Discriminability Sequence
D. NO-KR, Low to High Discriminability Sequence

Table 4. Contrasts Between Condition Means in Experiment 1

Contrast	Conditions	Differences in d'	ß	|ß-ß*|	ßVAR
A	KR - No KR	0.47 *	-2.03	-1.86 *	-0.026
B	[Lo to Hi] [Hi to Lo]	-0.38 *	-0.90	-0.18	-0.015
C	High - Low	1.65 ***	2.33 ***	2.21 ***	0.020
D	0.2 - 0.4	0.13	-0.92	-0.08	0.004

* $p<.05$ ** $p<.01$ *** $p<.001$

NOTE:
A. KR
B. Discriminability Sequence
C. Discriminability
D. Defect Probability

Probability factor did not attain significance in d' ($F[1,48]=2.20$, $p>.1$). Among these independent manipulations, there were two significant interactions for sensitivity. The KR X Defect Discriminability interaction ($F[1,48]=7.67$, $p<.01$) showed that KR was more effective in increasing sensitivity for High compared to Low Discriminability defects. KR also increased sensitivity more in the 0.4 defect probability condition relative to the 0.2 condition as evidenced by the significant KR X Probability interaction ($F[1,48]=5.05$, $p<.05$).

Response Criterion (ß)

Defect discriminability was the only significant main effect on inspector response criterion, with High Discriminability defects increasing ß by an average of 2.33 (see Table 3) compared to Low Discriminability defects ($F[1,48]=16.10$, $p<.001$). Although neither the KR nor the defect probability main effects significantly influenced ß ($p>.05$), there was a significant interaction between these two factors ($F[1,48]=4.63$, $p<.05$). ß was lowered from 5.9 to under 2.3 for KR in the 0.4 defect probability condition. This interaction is clearly shown in Figure 7. A second interaction was also observed between Defect Discriminability Sequence and Defect Probability ($F[1,48]=6.42$, $p<.05$). This interaction, also shown in Figure 7, is complex in its interpretation: as defect probability increased from 0.2 to 0.4, those who started in the Low discriminability condition became more conservative (larger ß) than those who started in the High discriminability condition.

(a) (b)

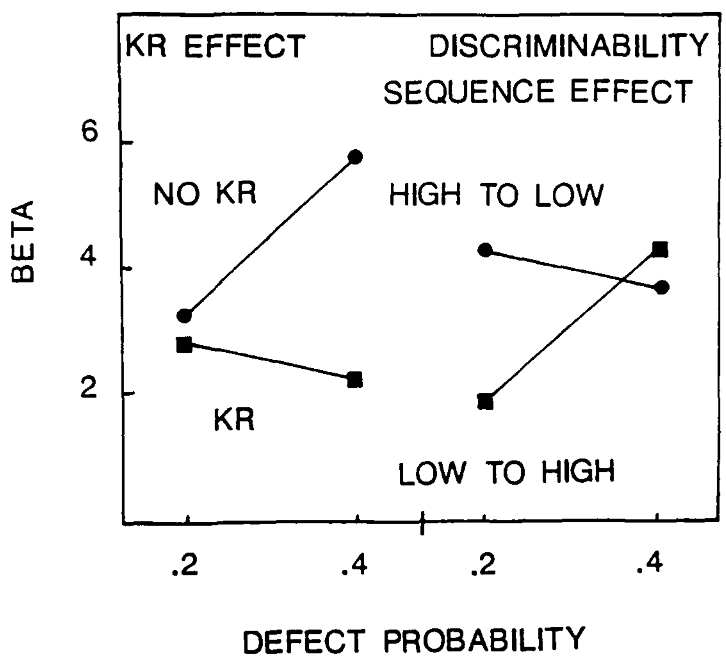


Figure 7. ß Interactions for Defect Probability. (a) KR X Defect Probability Interaction. (b) Discriminability Sequence X Defect Probability Interaction

A third interaction was also observed between Defect Discriminability Sequence and the level of Defect Discriminability ($F[1,48]=6.06$, $p<.05$) with ß being dramatically lowered for High discriminability defects in the Low to High Discriminability sequence compared to the High to Low sequence.

Optimal Response Criterion (ß*)

Relative success in shifting one's ß to its optimal value is an important indication of training progress. A ß optimality score was computed from $|ß-ß^*|$, with smaller values associated with more optimal performance. ß* for defect probabilities 0.2 and 0.4 were computed as 4.0 and 1.5, respectively, from Equation 3. ANOVA for $|ß-ß^*|$ showed significant main effects of KR ($F[1,16]=6.66$, $p<.05$) and Defect Discriminability ($F[1,48]=15.13$, $p<.001$). Both KR and Low discriminability defects produced significantly lower $|ß-ß^*|$ than NO-KR and High Discriminability defects. There were also significant KR X Defect Discriminability ($F[1,48]=4.75$, $p<.05$) and KR X Defect Probability ($F[1,48]=6.19$, $p<.05$) interactions, with the KR advantage more pronounced for High Discriminability defects and higher defect probabilities, as shown in Figure 8.

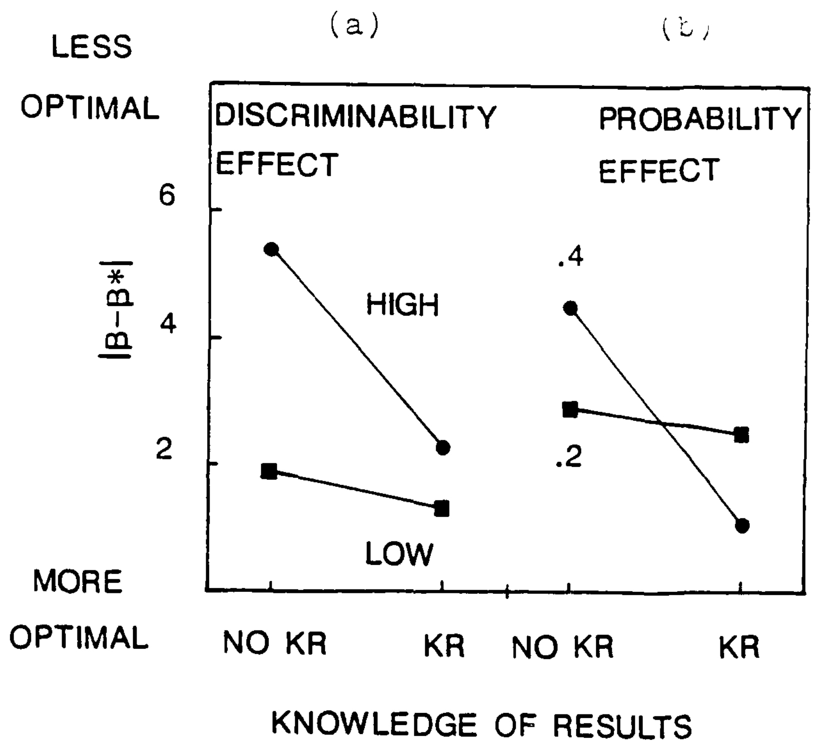


Figure 8. |ß-ß*| Interactions for KR. (a) Discriminability X KR Interaction. (b) Defect Probability X KR Interaction.

Relationship Between Sensitivity and Criterion

The relationship between d' and ß was evaluated to determine if any systematic dependence existed between these two parameters. A multiple regression for d', using the four independent variables plus ß as predictor variables, illustrated that the coefficient for ß was not significant ($t(df=1) = -0.77$, $p > .1$). A correlation analysis was also performed to determine if subjects with greater sensitivity (higher d' values) were also able to shift their ß's more optimally (lower $|ß - ß^*|$). The correlation between $|ß - ß^*|$ and d' was only 0.164, which was not significance ($F[1,78] = 2.15$, $p > .1$).

ß Variability

The ANOVA showed that no main effect was significant for ßVAR (see Tables 3 and 4). In particular, KR clearly did not significantly reduce the variability in inspectors' response criterion. The only significant interactions included the Defect Discriminability Sequence X Defect Discriminability Level term ($F[1,48] = 14.61$, $p < .0001$) and a three-way term which included the above two variables and Defect Probability ($F[1,48] = 5.72$, $p < .05$).

Regression analysis provided no evidence that a negative linear relationship existed between d' and ßVAR within any condition. Higher values of d' were not associated with reduced ßVAR. Sensitivity and ß variability were negatively correlated for subjects in the KR/High to Low Defect Discriminability Sequence (r= -0.257), however the regression was not significant ($F[1,18]=1.27$, $p>.1$). There was a significant positive linear relationship between d' and ßVAR ($F[1,18]=12.16$, $p<.01$) for subjects in the KR/Low to High Defect Discriminability Sequence condition. In summary, ßVAR was not a significant predictor of d' values ($t=1.11$, $p>.1$).

Reaction Time

As accuracy is usually correlated with observation time in visual detection studies, a separate analysis of Reaction Time (RT) was performed. Defect Discriminability ($F[1,48]=24.33$, $p<.001$) was the only significant main effect, with lower RT's for High Discriminability defects. Although mean RT was 300 msec faster for the KR condition, this difference was not significant due to high between-subject variability. The presence of a significant Discriminability Sequence X Defect Discriminability interaction ($F[1,64]=4.28$, $p<.05$) indicated that decreasing defect discriminability resulted in a greater increase in RT for those who started with Low discriminability defects than for those who started with High discriminability defects. There was also a KR X Defect Discriminability interaction ($F[1,48]=9.17$, $p<.01$) showing that KR was effective in significantly lowered reaction times in the Low Discriminability condition.

Discussion

This experiment demonstrated that KR significantly increased sensitivity and reduced the amount of time required to make a decision in a visual inspection task. The increase in sensitivity was more pronounced for High Discriminability defects and 0.40 defect probability while the faster reaction times were associated with Low Discriminability defects. The overall effect of KR was mediated by an increase in HR, as opposed to a significant decrease in FAR, likely acting as an additional source of information for increasing one's defect knowledge and allowing better discriminability between defects and nondefects during training. Some investigators, however, have argued that the primary effect of KR is motivational rather than informative (Weidenfeller, Baker, and Ware, 1962; Antonelli and Karas, 1967). Observed collateral effects, here, suggested that KR is providing a specific type of information which can be used to improve defect detection performance.

The present relationship between KR and response criterion casts some doubt on the purely motivational aspects of performance feedback. According to theory underlying β^*, as the defect probability increases, one's β should decrease, concomitant with more liberal responding. Here, KR failed to significantly lower β as defect probabilities increased from 0.2 to 0.4. Therefore, subjects may not have extracted the necessary defect distribution information from KR to lower β. KR did, however, move β towards its optimal value, especially in the 0.4 defect probability condition. The effect of KR on β in this instance certainly appeared to reverse the extreme nonoptimality of the No-KR condition, although it still did not precisely follow the normative predictions of the β^* model. If effects are assumed to be primarily motivational, KR should produce similar effects on both d' and β. The disparate effects of KR on d' and β suggested that specific information was provided which enhanced sensitivity but had little effect on the magnitude or variability of response criterion. While increased motivation may be an important part of KR, it may only come about as a result of the superior performance achieved through greater defect knowledge. According to this view, higher performance causes increased motivation rather than vice versa.

An implication from the overall failure of KR to influence β is the existence of an inherent difference in representing defect specification versus distribution knowledge in the cognitive system. The physical characteristics of defects may be more easily extracted, stored and accessed at a later time than associated probabilistic information. Mental or internal models may also be used to accentuate this advantage. Probabilistic judgement, on the other hand, likely relies more on heuristics and subjective biases of the human decision maker (Kahneman, Slovic, and Tversky, 1986). These less structured, informal rules likely require more development time and less competition from other aspects of the inspection task. Despite the inability of KR to produce optimal β shifts as defect probabilities increased, the presence of KR clearly resulted in more overall optimal criterion placement, as measured by $|\beta-\beta^*|$, and dramatically reversed the extreme conservatism of NO-KR subjects in the 0.40 probability condition. One way to account for these results is to view defect distribution knowledge as consisting of several different types; e.g., knowledge of defect frequencies and knowledge of criterion placement. This idea is consistent with Embrey's (1975) view that the ability to estimate the probability of a defect is a completely separate attribute from one's skill in optimally adjusting ß. The effect of KR on defect distribution knowledge in this experiment is positive to the extent that it provides frequency information to inspectors, which reduces the extreme conservatism present in the NO-KR condition. KR, however, failed to translate this knowledge into optimal ß shifts. The net result was an overall tendency toward optimality due to KR without necessarily following the predictions of the $ß^*$ model.

Another possible explanation for the higher inspector sensitivity observed with KR is that KR provides the inspector with response criterion knowledge which is then used to reduce the variability of ß and increase the effective sensitivity (Drury, 1988). However, the results of this experiment showed that KR did not significantly reduce ß variability nor was d' related to the ß variability measure in any of the experimental conditions. Thus, KR increased sensitivity by enhancing defect knowledge.

Defect discriminability, defined by the difference between the mean lengths of defective and non-defective line segments, had a strong influence on both d' and ß. Decreasing the discriminability reduced HR while increasing both FAR and RT. Changes in sensitivity, here, were not related in any systematic way to changes in response criterion, so high sensitivity does not necessarily allow one to perform closer to ß*. Thus, the abilities which are measured by d' and ß are independent and require different training strategies.

The Sequence of Defect Discriminability levels proved to be a significant predictor of sensitivity, in that starting with a higher difficulty task allowed subjects to maintain a higher d' than those starting with a lower difficulty task. The advantage of a high to low difficulty (Low to High Discriminability) sequence held up for both difficulty levels. One interpretation of this effect is in terms of the mental workload demands of the inspection task. Lintern and Wickens (1987) used attention theory to explain the effect of mental workload on skill acquisition and task training. According to a resource view of attention, skilled task performance requires the investment of limited processing resources, which must be allocated in greater amounts as the demands of the task increase. Mane and

Wickens (1986) predicted that increasing the mental workload of a task to be trained should result in better learning if the source of this increased load directly benefits task learning (intrinsic task component). On the other hand, if higher workload stems from aspects of the task which do not directly benefit the target of learning (extrinsic task component), then learning should deteriorate and performance should be lowered.

Applying a resource theory framework to the present inspection task can help explain the greater sensitivity observed in the High to Low Difficulty (Low to High Discriminability) sequence subjects. For subjects who started out in the High Difficulty condition, the greater mental workload associated with the Low Discriminability defects forced these subjects to invest more resources in learning the critical characteristics of the defects. As a result, when next performing the Low Difficulty task, their enhanced sensitivity due to better learning produced significantly higher d' values than those who started with the Low Difficulty task. Performing the Low Difficulty task first may have failed to motivate subjects to invest sufficient processing resources to learn the finer details of defects. In fact, sensitivity was higher when the high difficulty task was performed first than when it was performed after "practicing" on the Low Difficulty task.

The failure of the High to Low Difficulty Sequence to similarly enhance the setting of a more optimal inspector response criterion may also reflect the influence of mental workload demands on task learning. In this case, however, the higher workload of the Low Discriminability defects was extrinsic to the task of learning the defect probability distribution necessary to optimally adjust ß. Therefore, shifting ß optimally in response to increasing defect probabilities was inhibited during learning by the higher workload demands imposed by an extrinsic aspect of the inspection task. This distinction between intrinsic and extrinsic task components can be used as a basis for deciding which components should be trained as well as their relative difficulty levels.

Chapter 4

EXPERIMENT 2: THE EFFECTS OF FALSE KR AND DECISION PAYOFFS ON VISUAL INSPECTION PERFORMANCE

The overall objectives of this experiment was to 1. compare the relative effectiveness of FALSE KR versus TRUE KR for increasing inspector sensitivity; and 2. determine the effects of changing decision payoffs on the optimization of inspector response criterion.

Method

Subjects

Eighteen right-handed male volunteers from an introductory Human Factors course were recruited to participate in this experiment. Each was screened for 20/20 or better corrected visual acuity. Payment of up to $5.00/hour was made; this included a bonus payment for correct responses. The experiment lasted about 1.5 hours.

Apparatus

The equipment used in this experiment was identical to that used in Experiment 1.

Experimental Design

Subjects were randomly assigned to one of three feedback groups: NO-KR, TRUE-KR, or FALSE-KR. All subjects performed the LOW discriminability version of the inspection task described in Experiment 1 in conjunction with three independent variables schematized in Table 5 and described below.

Knowledge of Results

The presentation of KR for the NO-KR and TRUE-KR groups was exactly the same as the presentation for NO-KR and KR groups of Experiment 1. In the FALSE-KR group, however, subjects received incorrect feedback on selected trials. Data from pilot testing established when inspectors received False KR to prevent subjects from quitting the experiment. If the length of a line segment was between .48 and .52 of the width of the viewing area, then KR ("right" or "wrong") was randomly presented to the subject at the top of the

Table 5. Experiment 2 design.

DECISION PAYOFFS	DEFECT PROB.	NO-KR	TRUE-KR	FALSE KR
SYMMETRIC	0.2			
	0.4	Block I	Block II	Block III
HR 2X	0.2			
		Ss= 1-6	Ss= 7-12	Ss= 13-18
	0.4			
HR 9X	0.2			
	0.4			

NOTE: Sequence of Payoff conditions counterbalanced between subjects.

screen. Summary false accuracy and monetary reward information was also provided after each 80-trial block. Instructions in the two KR conditions specified that the number of defects could change between blocks; no such information was given in the NO-KR condition.

Decision Payoffs

Three different payoff conditions were used within each subject. For each payoff condition, the values/costs of various decision outcomes were verbally communicated to the subject at the beginning of a block of trials. In the Symmetric payoff condition, the values and costs of all decisions were equal. In the 2X condition, the value of a hit and the cost of a miss were twice as high as the value of a correct acceptance and the cost of a false alarm. In other words, there was a modest financial gain for subjects who maximized correct detection of defects (hits) and minimized misses. In the 9X condition, the value of a hit and the cost of a miss were 9 times as high as the other two decision outcomes. Subjects now received a substantial financial reward for maximizing hits and minimizing missed defects.

Defect Probability

The number of defects was manipulated within each subject from 16 defects per 80 trials for the 0.2 probability condition to 32 defects per 80 for the 0.4 condition.

Each subject performed a total of six blocks of 80 inspection trials. The blocks were divided into three pairs, one pair for each payoff condition. The first block of each pair was presented at defect probability 0.2 and the second at 0.4, replicated across the three payoff conditions. The sequence of payoff conditions was counterbalanced across subjects.

Subjects were nested within KR levels while both Payoff and Defect Probability were within subject variables. All main effects are derived from fixed factors with the subject factor randomized (see Appendix A for a complete description of the statistical model).

Procedure

Prior to performing the task, each subject observed 20 inspection trials presented by the experimenter with the halfway criterion in place. The difference between defective and nondefective line segments was carefully explained and reinforced. Each subject practiced the task for four blocks of 25 trials, two blocks at defect probability 0.2 and two blocks at 0.4. Performance of at least 90% correct decisions was required on the last two blocks to be admitted into the experimental phase. After training, each subject performed six blocks of 80 trials according to the KR group and payoff sequence assigned. Two minutes rest was provided between blocks. HR, FAR, and RT were recorded for each block. Derived values of d', ß, $|ß-ß^*|$ were also calculated for each condition. A bonus system, similar to the one used in Experiment 1, reinforced high inspection performance for a given payoff condition.

Results

Reported statistics below are from three-factor ANOVA's on each dependent variable. Mean values for dependent measures are included in Table 6. Table 7 contains F values for main effects along with error probabilities. Pairwise comparisons between significant treatment means were made using Duncan's Multiple Range Test. Appendix B shows that the assumptions of the ANOVA model were met. In addition, Appendix B also shows that the SDT assumptions of normality and equal variance were generally fulfilled.

Hit Rate

All main effects were significant for HR. Inspectors in the NO-KR condition had significantly lower HR's than those in either the TRUE-KR or FALSE-KR conditions ($F(2,15)=10.53$, $p<.01$). There was no difference between TRUE-KR or FALSE-KR inspectors. HR was also the lowest in the Symmetric Payoff condition ($F(2,75)=9.69$, $p<.001$) and in the 0.4 Defect Probability condition ($F(1,75)=7.59$, $p<.01$). In other words, inspectors correctly detected defects more often in the more liberal payoff conditions and when defect probability was low. No interactions were present.

Table 6. Mean values for d', ß, and |ß-ß*| in Experiment 2

	DEFECT PROB.	NO-KR			TRUE-KR			FALSE KR		
		d'	ß	|ß-ß*|	d'	ß	|ß-ß*|	d'	ß	|ß-ß*|
A	0.2	1.75	1.60	2.40	2.42	2.84	1.87	1.91	0.68	3.32
	0.4	1.67	4.48	3.04	1.90	1.89	0.43	2.20	1.90	1.13
B	0.2	1.81	1.13	0.93	2.31	1.12	0.95	1.95	0.44	1.56
	0.4	1.55	3.04	2.29	2.35	1.50	0.96	1.68	0.67	0.17
C	0.2	1.70	0.82	0.44	2.12	1.50	1.06	1.58	0.49	0.22
	0.4	1.82	1.81	1.65	2.20	1.22	1.07	1.74	0.35	0.19

NOTE: A. Symmetric decision payoff
B. Hits 2X more valuable than false alarms
C. Hits 9X more valuable than false alarms

Table 7. Calculated F-values for Main Effects in Experiment 2

Main Effect	d'	ß	|ß-ß*|	RT
KR	5.50*	4.77*	4.40*	0.64
PAYOFF	1.33	10.35***	10.16***	0.33
DEFECT PROBABILITY	0.04	9.56**	0.75	0.17

* $p<.05$ ** $p<.01$ *** $p<.001$

False Alarm Rate

Only KR and Payoff main effects on FAR were present. FALSE-KR inspectors had the highest FAR with no difference between NO-KR and TRUE-KR inspectors ($F(2,15)=11.38$, $p<.01$). Inaccurate performance information apparently caused inspectors to make more errors identifying defects than no information at all. FAR significantly increased from the Symmetrical Payoff condition through 2X and 9X conditions as inspection instructions became increasingly more liberal ($F(2,75)=31.2$, $p<.001$). Inspectors were more likely to incorrectly identify defects as the value of a hit and the cost of a miss increased. No interactions were present.

Sensitivity (d')

The only main effect present for d' was KR. TRUE-KR inspectors had significantly higher sensitivity than either NO-KR or False-KR inspectors. On the average, inspectors who received accurate information about their performance increased their sensitivity by over 20% compared to inspectors receiving FALSE-KR and over 25% for NO-KR. This advantage of TRUE-KR was consistent across all experimental conditions. No interactions were present.

Response Criterion (ß)

All main effects were significant for ß. Both KR groups had lower ß's than NO-KR with FALSE-KR being significantly lower ($F(2,15)=4.77$, $p<.05$). Inspectors receiving inaccurate performance information had more liberal response criteria than either NO-KR or TRUE-KR inspectors.

Payoffs also produced significant changes in ß as the 2X and 9X conditions, which emphasized the value of hits over false alarms, lowered ß compared to the Symmetrical Payoff condition ($F(2,75)=10.35$, $p<.001$). Lowering ß as the value of a hit (and cost of a miss) increased is in accordance with the SDT model.

On the other hand, increases in ß as defect probabilities increased from 0.2 to 0.4 ($F(1,75)=9.56$, $p<.01$) violated this model. This discrepancy can be explained by considering the KR X Defect Probability interaction ($F(2,75)=8.39$, $p<.01$) shown in Figure 9. In both the NO-KR and FALSE-KR conditions, ß significantly increased as defect probabilities increased; however, TRUE-KR inspectors decreased ß as predicted.

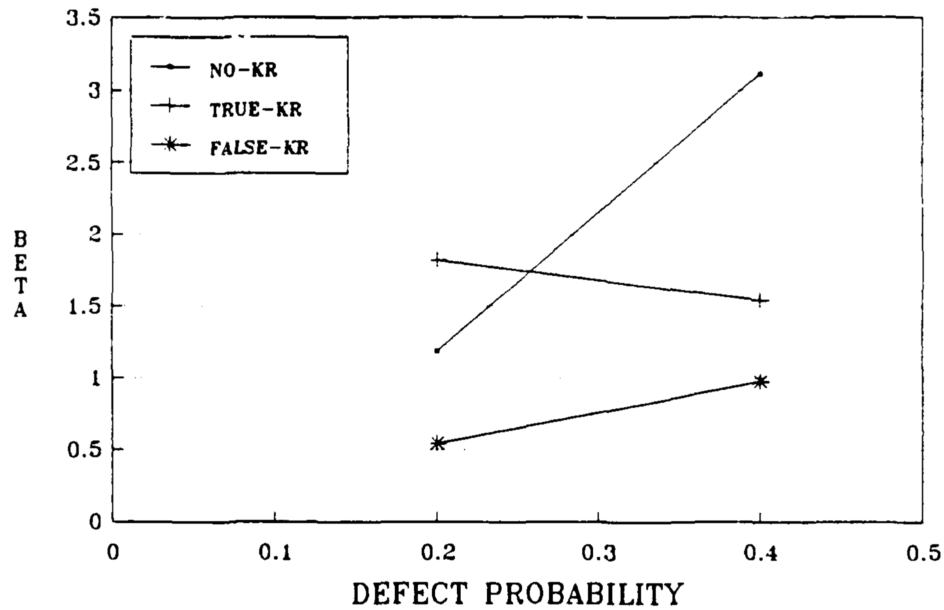


Figure 9. Defect Probability X KR Interaction for ß

The consistent advantage of TRUE-KR in manipulating ß between both payoff conditions and defect probabilities is clearly illustrated in Figure 10. Each Block represented a specific Payoff/Defect Probability combination.

1. Block 1 - Symmetrical/0.2
2. Block 2 - Symmetrical/0.4
3. Block 3 - 2X/0.2
4. Block 4 - 2X/0.4
5. Block 5 - 9X/0.2
6. Block 6 - 9X/0.4

Duncan's Multiple Range Test (Montgomery, 1984) for pairwise comparisons of mean ß's within each block against ß* showed no significant difference between TRUE-KR ß points and ß* points across all blocks. Figure 10 shows that NO-KR inspectors had trouble manipulating ß in the 0.4 condition while FALSE-KR inspectors had trouble in the 0.2 condition.

Optimality Scores (|ß-ß*|)

Inspectors in the NO-KR condition had significantly higher scores (less optimal) than inspectors in either KR condition ($F(2,15)=4.4$, $p<.05$). Both TRUE and FALSE KR resulted in more optimal performance. In addition, overall

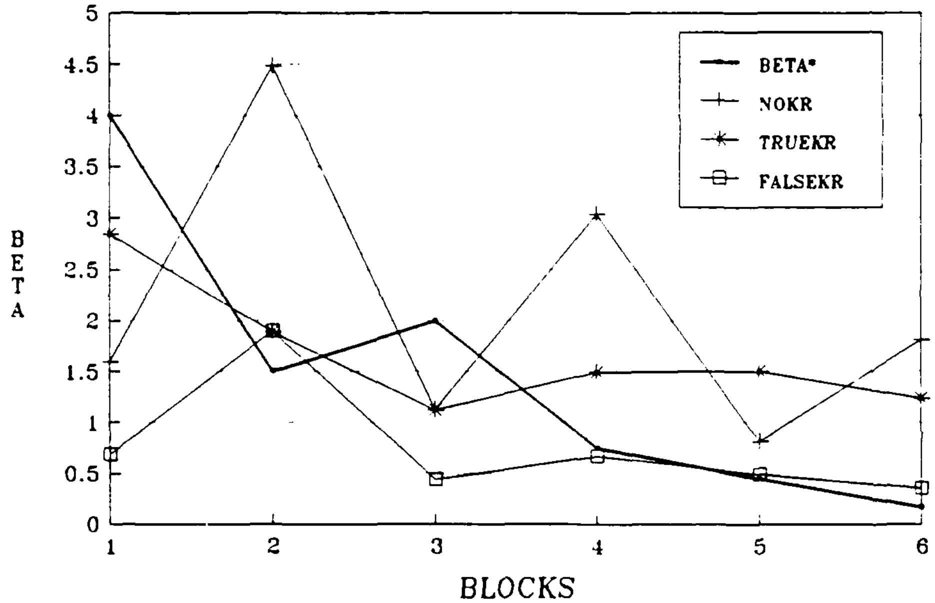


Figure 10. Payoff X KR Interaction for ß

inspector performance in the Symmetric Payoff condition was significantly less optimal than the other two conditions ($F(2,75)=8.12$, $p<.001$). While there was no significant difference in scores for the two probability conditions ($F(1,75)=.75$, $p>.1$), there was an interaction between KR and Defect Probability ($F(2,75)=8.12$, $p<.01$). Inspectors receiving NO-KR were less optimal as defect probabilities increased while both KR groups became more optimal.

Reaction Time (RT)

There were no significant main or interactive effects of any of the independent variables on RT. However, RT was 17% lower, on the average, for KR inspectors, although this decrease was not significant ($F(2,15)=.64$, $p<.1$).

Discussion

The experimental results obtained thus far clearly demonstrated that KR significantly increased inspector sensitivity and resulted in overall more optimal criterion placement than NO-KR inspectors. In addition, there was substantial evidence that KR also reduced the time to make an inspection decision without sacrificing accuracy. In

Experiment 1, the increase in sensitivity was more pronounced for High Discriminability defects and 0.4 defect probability while faster RT's were associated with Low Discriminability defects. The overall effect of KR in both experiments was mediated by a significant increase in HR, as opposed to a decrease in the FAR, likely acting as an additional source of information for increasing one's defect knowledge and allowing better discriminability between defects and nondefects.

While the ability of KR to increase inspector sensitivity in Experiment 1 could be attributed to either motivational or informational aspects of performance feedback, the results of Experiment 2 clearly supported the informational explanation. Subjects who received TRUE-KR had significantly higher d's than subjects who received either FALSE-KR or NO-KR (with no significant difference between the latter two groups). If increased motivation was responsible for the higher sensitivity than there should have been no difference between the TRUE-KR and FALSE-KR groups. Also, since FALSE-KR inspectors only received incorrect feedback on selected trials where the distinction between defects and nondefects was more problematic, it is apparent that the more difficult trials are important for acquiring higher levels of defect knowledge. Although increased motivation may certainly be an important aspect of any KR technique, it may only come about as a result of better performance achieved through greater defect knowledge.

The effects of KR on inspector response criterion are less clear cut but still encouraging. According to the theory underlying ß*, as defect probability increases, one's ß should decrease, concomitant with more liberal responding. In Experiment 1, KR did not significantly lower ß as defect probabilities increased from 0.2 to 0.4. Apparently, subjects did not extract the necessary defect distribution information from KR to optimize beta. However, KR did shift ß toward its optimal value, especially in the 0.4 defect probability condition. This result was closely replicated in Experiment 2 where TRUE-KR did not change ß significantly as a function of defect probability but did move ß closer to optimal then either FALSE-KR or NO-KR. The effect of KR on ß, in these instances, reversed the extreme nonoptimality of the NO-KR condition, although it still did not precisely follow the normative predictions of the ß* model.

If the defect probability and payoff variables manipulated in Experiment 2 are combined, then a comparison can be made between ß and KR as a function of the six Defect Probability X Payoff conditions. The result showed that

TRUE-KR tracked the β^* model more closely than either of the other KR groups. For NO-KR, 4 of the 6 β data points were significantly different from the corresponding β^* points for that particular condition. For FALSE-KR, 2 of the 6 were significantly different while for TRUE-KR, none of the β data points were significantly different from β^*. Therefore, TRUE-KR inspectors were able to manipulate their response criterion more optimally as both payoffs and defect probabilities changed during the task.

The results of the $|\beta-\beta^*|$ scores generally supported the superiority of TRUE-KR. Across all conditions, both TRUE-KR and FALSE-KR resulted in significantly lower (more optimal) scores than NO-KR. TRUE-KR was particularly effective in producing lower scores for more conservative (higher) β^*'s while FALSE-KR was associated with lower scores for more liberal (lower) β^*'s.

The cognitive model of KR utilization presented earlier should be modified in light of the current results. First, the primary advantage of KR was in increasing inspector sensitivity by enhancing defect knowledge. While both KR groups had more optimal β placement, overall β performance did not follow the optimal model, especially when manipulated by defect probabilities. Second, the concept of defect distribution knowledge should be broken down into two components based on the experimental evidence: Defect frequency knowledge and criterion placement knowledge. While KR provided defect frequency knowledge, the knowledge to translate this to the actual placement of one's response criterion may require more specific and detailed information.

Chapter 5

EXPERIMENT 3: TRAINING FOR DECISION MAKING IN VISUAL INSPECTION: INFLUENCE OF TASK DIFFICULTY AND KR

The objectives of this experiment were: 1. to determine if inspectors trained with KR maintained their higher sensitivity when KR was removed; 2. to determine if inspectors trained on high difficulty defects had higher sensitivity when subsequently performing the Low difficulty task; and 3. to evaluate the effects of training on retention of visual inspection skill.

Method

Subjects

Twenty right-handed males were recruited from a local newspaper ad for this experiment. Each was screened for 20/20 or better corrected visual acuity. Payment of up to $5.00 /hour was made; this included a bonus payment for superior performance. Total experimental time was 2 hours over two sessions.

Apparatus

The equipment in this experiment was identical to that used in the previous two experiments.

Experimental Design

Each subject was randomly assigned to one of four training groups in conjunction with three independent variables schematized in Table 8 and described below.

Inspector Training

Subjects were divided into four groups based on the level of KR (NO-KR, KR) and task difficulty (Low, High):

1. Group I - NO-KR/Low Difficulty
2. Group II - NO-KR/High Difficulty
3. Group III - KR/Low Difficulty
4. Group IV - KR/High Difficulty

The NO-KR, KR groups used here were treated exactly the same as in Experiment 1. Also, the difficulty levels correspond to the defect discriminability levels described in Experiment 1; low difficulty was characterized by High Discriminability defects and high difficulty was characterized by Low Discriminability defects.

Table 8. Experiment 3 Design

		TRAINING GROUPS							
		GROUP I		GROUP II		GROUP III		GROUP IV	
		(REPLICATIONS #1 & #2)							
PHASE	DEFECT PROB	#1	#2	#1	#2	#1	#2	#1	#2
1	0.2								
	0.4								
		BLOCK 1		BLOCK 2		BLOCK 3		BLOCK 4	
2	0.2								
		Ss= 1-5		Ss= 6-10		Ss= 11-15		Ss= 16-20	
	0.4								
3	0.2								
	0.4								

NOTE: Group I - NO-KR/Low Difficulty
Group II - NO-KR/High Difficulty
Group III - KR/Low Difficulty
Group IV - KR/High Difficulty
All subjects within a block performed two replications at each probability level for all three phases.

Experimental Phase

The experiment was divided into three separate phases:

1. Phase 1 - Training (based on group assignment)

2. Phase 2 - Immediate posttest (No-KR/Low Difficulty)

3. Phase 3 - Repeat Phase 2 (three weeks later)

Each phase was further divided into four blocks of 100 trials each. Phases 1 and 2 were performed on the same day while Phase 3 was performed 3 weeks later. Training Group manipulations were present only during Phase 1. During phases 2 and 3, a standard NO-KR/Low difficulty was presented regardless of the training group assigned. Performance on this "standard" task was used to evaluate inspector training.

Defect Probability

Just as in the previous two experiments, the number of defects varied within a phase from 20 for the 0.2 condition, to 40 for the 0.4 condition. Each subject always performed two replications of the 0.2 condition first followed by two replications of the 0.4 condition across all 3 phases.

Five subjects were randomly assigned to each training group. During Phase 1, each group performed their assigned training task for two blocks of trials at 0.2 defect probability followed by two blocks at 0.4. During Phase 2, all groups switched to the NO-KR/Low Difficulty task for the same block sequence as Phase 1. Subjects then repeated the Phase 2 task three weeks later for Phase 3 (see Appendix A for the appropriate statistical model).

Procedure

Each subject initially observed 20 trials with the half-way criterion visible and held constant. Next, 25 trials with five defects were presented to familiarize each subject with the experimental equipment and method of responding. Minimal practice was given to avoid "pre-training" subjects. Phase 1 began immediately with four blocks of 100 trials with the assigned training task. The first two blocks were always at 0.2 defect probability and the last two at 0.4. A two-minute rest period was given between blocks and a five-minute rest period between Phases. In Phase 2, all subjects performed four blocks of 100 trials with a NO-KR/Low Difficulty version of the inspection task. Again, the first two blocks at 0.2 and the second two at 0.4 defect probability. Phase 2 inspection task was repeated three weeks later during Phase 3. HR, FAR, and RT were recorded for each block during each phase. Values for d', ß, |ß-ß*| were calculated. The same bonus system used in previous experiments was implemented during each phase.

Results

Table 9 contains mean values for HR, FAR, and RT while Table 10 contains mean values for d', ß, and |ß-ß*| for each training group by phase and defect probability. Reported statistics below are from three factor ANOVA's on each of these dependent variables. Appendix B shows that the assumptions of the ANOVA model were met. Duncan's Multiple Range Test was used for all pairwise comparisons of treatment means. Both inspector HR and FAR will be considered together in the following section since they are both used to derive the primary measures of inspection performance.

Hit Rate and False Alarm Rate

Although there was a small decrease in HR for Training Groups III and IV which were trained with KR, this change was not significant ($F(3,16)=1.34$, $p>.1$). In contrast, these groups had significantly less false alarms ($F(3,16)=8.70$, $p<.01$) than those groups trained without KR.

Table 9. Mean Values for HR, FAR, and RT in Experiment 3

		PHASE								
		1			2			3		
	DEFECT PROB	HR	FAR	RT	HR	FAR	RT	HR	FAR	RT
A	0.2	0.88	0.27	1.34	0.84	0.18	1.27	0.86	0.19	1.10
	0.4	0.84	0.15	1.26	0.84	0.16	1.11	0.87	0.14	0.93
B	0.2	0.81	0.35	0.93	0.91	0.26	0.76	0.93	0.27	0.76
	0.4	0.72	0.30	0.87	0.88	0.19	0.77	0.89	0.17	0.72
C	0.2	0.80	0.05	0.61	0.87	0.10	0.71	0.87	0.13	0.75
	0.4	0.87	0.05	0.62	0.77	0.09	0.70	0.73	0.09	0.75
D	0.2	0.66	0.11	0.79	0.91	0.09	0.79	0.89	0.07	0.80
	0.4	0.66	0.11	0.79	0.69	0.05	0.81	0.74	0.04	0.80

NOTE: A. Group I NO-KR/Low Difficulty defects
B. Group II NO-KR/High Difficulty defects
C. Group III KR/Low Difficulty defects
D. Group IV KR/High Difficulty defects

Table 10. Mean values for d', ß, and |ß-ß*| in Experiment 3

	DEFECT PROB	Phase 1 d'	Phase 1 ß	Phase 1 |ß-ß*|	Phase 2 d'	Phase 2 ß	Phase 2 |ß-ß*|	Phase 3 d'	Phase 3 ß	Phase 3 |ß-ß*|
A	0.2	1.90	0.60	3.40	2.07	1.01	2.99	2.08	0.96	1.10
	0.4	2.25	1.70	1.33	2.18	1.78	1.47	2.49	1.61	1.39
B	0.2	1.37	0.67	3.34	2.08	0.61	3.39	2.16	0.47	3.54
	0.4	1.15	0.93	0.58	2.30	1.99	2.00	2.38	1.02	1.08
C	0.2	2.76	5.00	3.31	2.62	2.41	3.18	2.42	1.34	2.66
	0.4	2.83	2.76	1.50	2.37	3.84	3.01	2.31	3.57	2.60
D	0.2	1.76	2.13	1.86	2.88	1.59	2.41	2.95	2.67	2.42
	0.4	1.75	2.56	1.27	2.28	5.01	3.51	2.65	4.95	3.72

NOTE: A. Group I NO-KR/Low Difficulty defects
B. Group II NO-KR/High Difficulty defects
C. Group III KR/Low Difficulty defects
D. Group IV KR/High Difficulty defects

Phase and Defect Probability also had significant effects on both with the Phase 1 (training) resulting in lower HR ($F(2,200)= 11.75$, $p<.001$) and higher FAR ($F(2,200)=9.85$, $p<001$) while phases 2 and 3 remained unchanged. The 0.2 Defect Probability condition had both a higher HR ($F(1,200)$, $p<.001$) and a higher FAR ($F(1,200)=28.49$, $p<.001$) over all conditions.

Interactions between these independent variables can be used to further isolate and analyze their effects on HR and FAR. The Training X Phase interaction was significant for both HR ($F(6,200)= 6.60$, $p<.001$) and FAR ($F(6,200)= 7.67$, $p<.001$). Figure 11 shows that, during training, Groups II and IV had predictably lower HR than their Group I and III counterparts since these groups were trained on higher difficulty defects. However, during Phase 2 where all groups performed the same No KR/Low Difficulty inspection task, HR increased dramatically for Groups II and IV and remained constant during Phase 3, three weeks later. In particular, Group II inspectors had the highest HR's of all groups during Phases 2 and 3 despite being trained on High Difficulty defects. HR for Groups I and III remained constant across the 3 phases. For FAR, Groups I and II

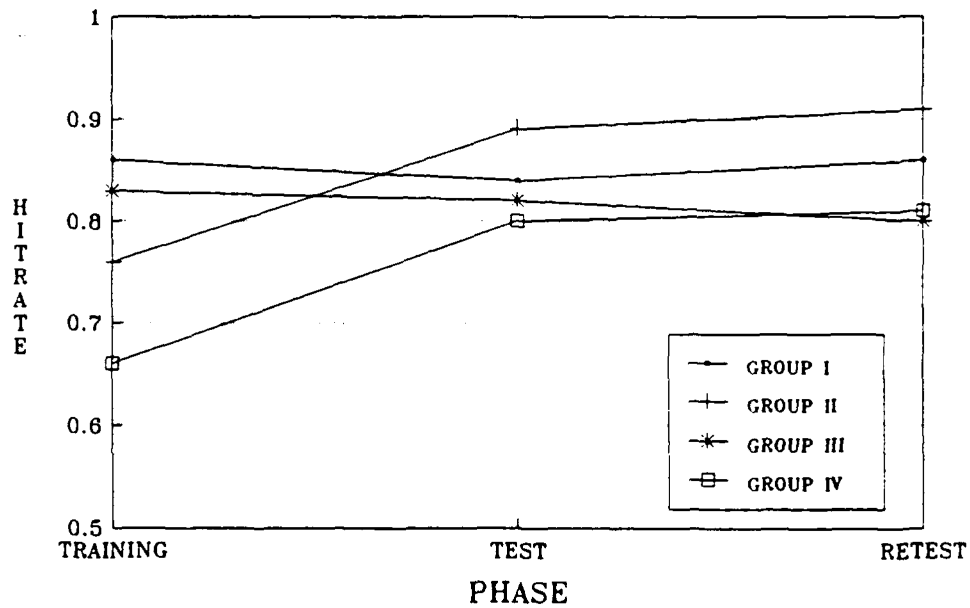


Figure 11. Phase X Training Group Interaction for Hit Rate

shown in Figure 12 had significantly higher FAR's during training which then decreased during phases 2 and 3. Inspectors trained with KR had significantly lower FAR's which remained constant across all 3 phases.

The Training X Defect Probability interaction was also significant for both HR ($F(3,200) = 3.78$, $p < .05$) and FAR ($F(3,200) = 3.37$, $p < .05$). Groups II, III, and IV all displayed a significant decrease in HR from 0.2 to 0.4 Defect Probability conditions. Group IV (KR/High Difficulty defect) showed the most dramatic decrease in HR while Group I (No KR/Low Difficulty defects) showed no change as a function of increasing probability levels. For FAR data, inspectors trained without KR (Groups I and II) had overall higher FAR which decreased as defect probability increased. In contrast, Inspectors trained with KR (Groups III and IV) had significantly lower FAR which remained constant across the defect probability conditions.

In addition, HR also exhibited a Phase X Defect Probability interaction ($F(2,200) = 3.29$, $p < .05$) where Phases

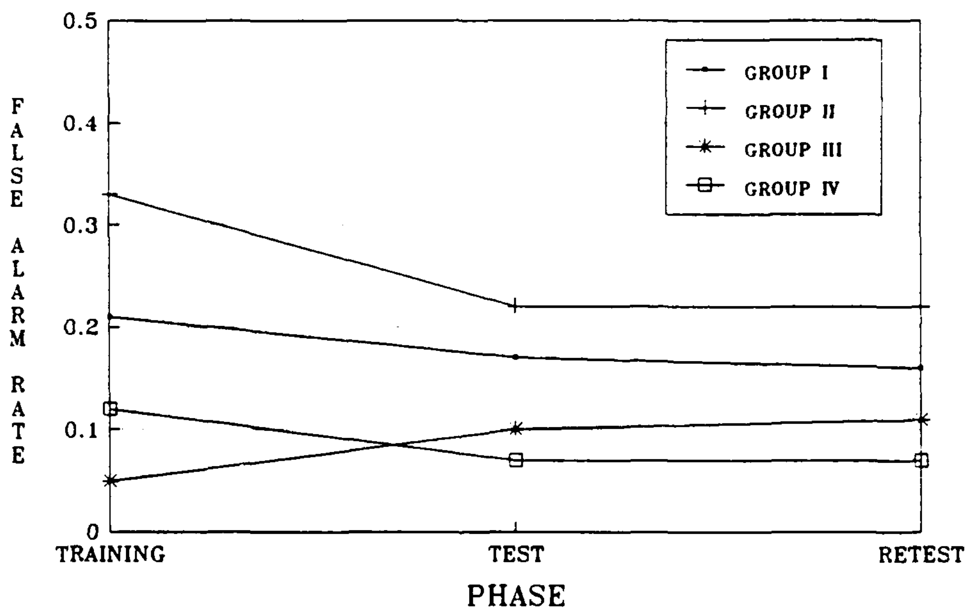


Figure 12. Phase X Training Group Interaction for False Alarm Rate

2 and 3 showed a significant decrease in HR as defect probabilities increased from 0.2 to 0.4. During Phase 1 (training), HR remains unchanged from 0.2 to 0.4 but at an overall lower level than Phases 2 and 3.

<u>Sensitivity (d')</u>

Inspector sensitivity as measured by d' was significantly affected by both Training Group and Phase. Inspectors trained with KR (Groups III and IV) had an overall higher mean d' ($F(3,16) = 3.79$, $p < .05$), especially on High Difficulty defects. Overall, Phase 1 (training) had a significantly lower mean d' than Phases 2 or 3.

The Training Group X Phase interaction ($F(6,200) = 18.95$, $p < .001$) illustrated in Figure 13 showed that inspectors trained on High Difficulty defects had expected lower d's during Phase 1, but the group trained with KR (Group IV) maintained a significant advantage during Phases 2 and 3 when the KR was removed. In contrast, Group III who also trained with KR but with Low Difficulty defects had the highest mean d' during training but than decreased during Phases 2 and 3. Groups trained without KR had essentially constant d's through phases 2 and 3.

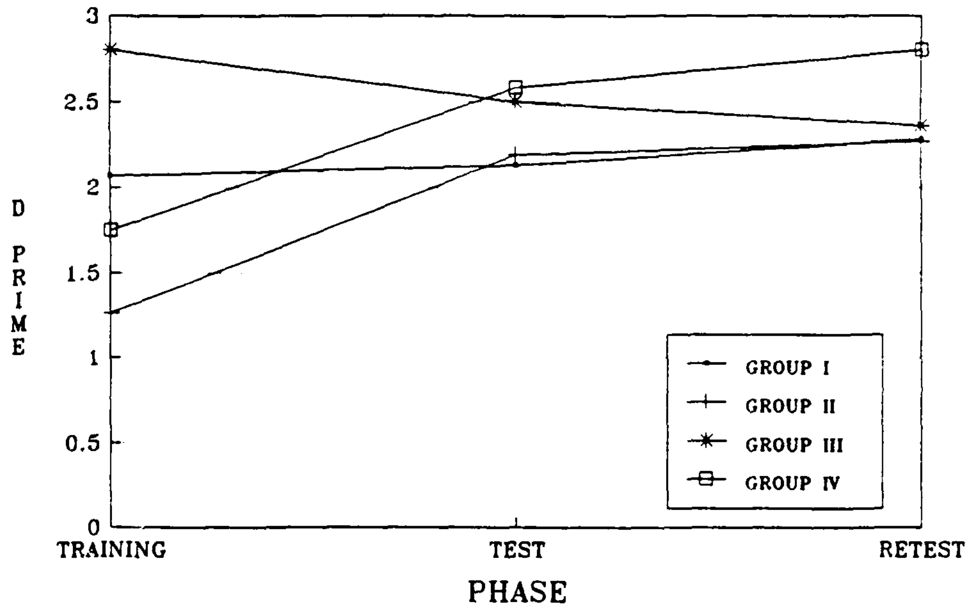


Figure 13. Phase X Training Group Interaction for d'

The Training Group X Defect Probability interaction further isolated the effects of the various training groups ($F(3,200)=6.14$, $p<.001$). Groups III and IV trained with KR had significantly higher d's than those trained without KR for the 0.2 Defect Probability condition. This difference becomes much smaller in the 0.4 condition where Group IV inspectors had significantly smaller d's from 0.2 to 0.4 while inspectors in Group I significantly increased their d's from 0.2 to 0.4. These trends were much more apparent for Phases 2 and 3 on the standard No KR/Low Difficulty inspection task.

<u>Response Criterion (ß)</u>

Inspector response criterion as measured by ß was significantly affected by both Training Group and Defect Probability. Inspectors in Groups III and IV had significantly higher ß's than Groups I and II ($F(3,16)=4.56$, $p<.05$). The presence of KR during training was associated with conservative decision making. In addition, going from 0.2 to 0.4 defect probabilities also produced significantly higher ß's ($F(1,200)= 17.13$, $p<.01$) for the 0.4 condition.

The Phase X Defect Probability interaction ($F(2,200)=19.8$, $p<.01$) showed that during Phase 1 (training) ß remained constant across probability conditions. However, there was a significant increase in ß for Phases 2 and 3 as defect probabilities increased from 0.2 to 0.4. This trend of conservative decision making with increasing defect probability was fairly consistent; the only exception was Group III inspectors trained with KR/Low Difficulty defects. Their performance during training depicted in Figure 14 reflected a trend of more liberal decision making (lower ß) as defect probabilities increased.

<u>ß Optimality Scores (|ß-ß*|)</u>

These scores were used to evaluate response criterion performance during inspection. Smaller deviations of ß from optimal ß or ß* are associated with higher inspector performance in terms of minimizing inspection error for a given sensitivity. While Training Group did not have a significant effect on these optimality scores ($F(3,16)=2.46$, $p>.1$), both Phase and Defect Probability resulted in significant changes. Inspectors were more optimal during Phase 1 (training) ($F(2,200)=3.51$, $p<.05$) and in the 0.4 defect probability condition ($F(1,200)=22.45$, $p<.001$). Optimality scores were lower overall during inspector training and high defect probability conditions.

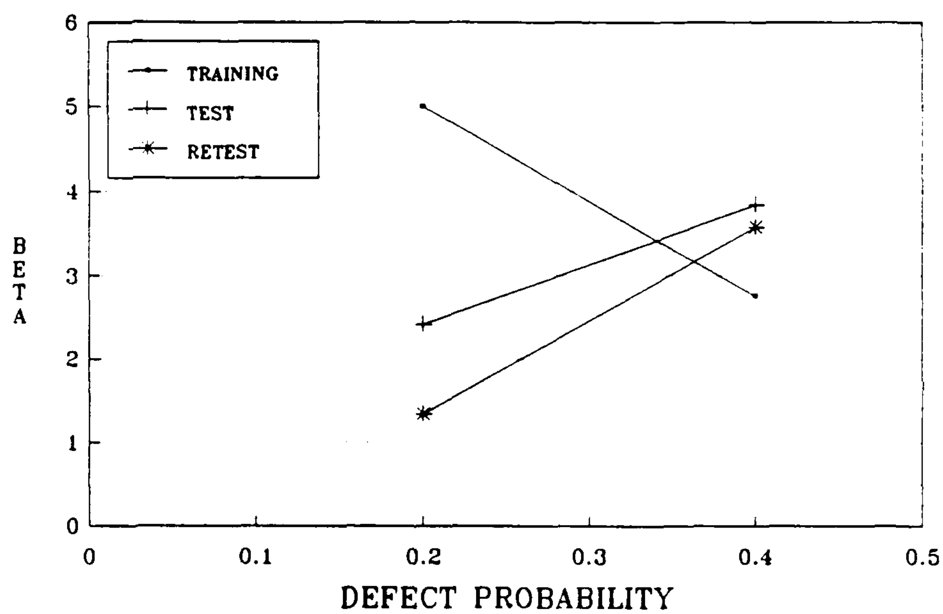


Figure 14. Defect Probability X Phase Interaction for Group III ß

The Training Group X Defect Probability interaction ($F(3,200)=8.61$, $p<.001$) showed that for Groups I and II (No KR) inspectors became more optimal as they moved from 0.2 to 0.4 defect probability conditions while Groups III and IV remained the same. At least part of this "optimal" performance of NO-KR inspectors is due to the extremely low criterion adopted by these inspectors regardless of the defect probability level. The Phase X Defect Probability interaction ($F(2,200)=3,64$, $p<.05$) showed that when defect probability is increased from 0.2 to 0.4, inspectors in Phase 1 (training) were more optimal than in either Phase 2 or 3. These interactions indicated that the increased optimality experienced by inspectors in Groups I and II when moving from 0.2 to 0.4 occurred primarily during training.

Reaction Time (RT)

Inspector RT was also measured to determine decision time under the various experimental conditions. Mean times are located in Table 8. Training Group had a significant effect on RT ($F(3,16)=5.55$, $p<.01$); Group I inspectors had the slowest RT compared to any of the other groups. Groups II, III, and IV all had faster and similar RT's. Both Phase and Defect Probability also affected RT with

Phase 1 ($F(2,200)=5.38$, $p<.01$) and 0.2 Defect Probability ($F(1,200)=4.91$, $p<.05$) producing significantly slower RT's than the other conditions.

The Training Group X Phase interaction ($F(6,200)=9.34$, $p<.001$) shown in Figure 15 clarifies the above main effects by showing that during Phase 1 (training), inspectors trained with KR had significantly lower RT's than inspectors in the NO KR groups. This advantage for Group IV disappears during Phases 2 and 3. The Training Group X Defect Probability interaction ($F(3,200)=3.40$, $p<.05$) show that RT decreased from 0.2 to 0.4 conditions for Group I inspectors and remained the same for the other 3 groups.

Discussion

The results of this experiment clearly show that KR trained inspectors performed at a higher level, not only during training when KR was actually present, but also during subsequent phases when KR was no longer available. Inspectors trained with KR showed higher sensitivity, more optimal response criterion shifts and faster decision times independent of task difficulty. In particular, sensitivity was higher during Phase 2 when KR was withdrawn for inspectors trained with KR regardless of the difficulty

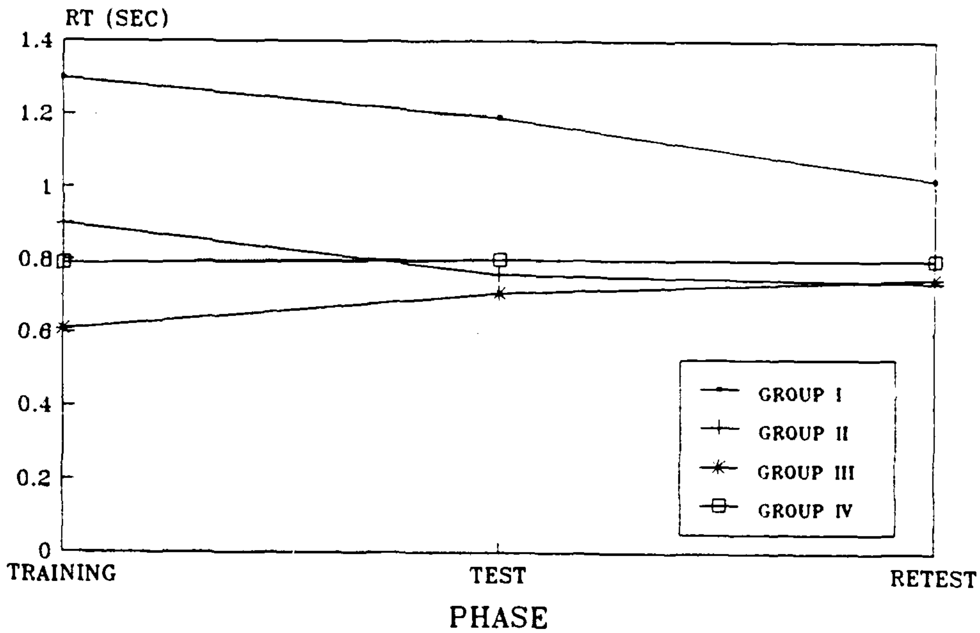


Figure 15. Phase X Training Group Interaction for RT

level. KR also continued to improve inspector performance during Phase 3, three weeks later. Clearly, a degree of permanence has been established for the superiority of KR during visual inspection.

Inspector sensitivity, in particular, was affected not only by the presence of KR during training but also by the degree of task difficulty. The superiority of KR trained inspectors was evident during the immediate post-test for both High and Low Difficulty conditions. Furthermore, in the three week retest (Phase 3) inspectors trained with both KR and High Difficulty defects continued to improve their performance, resulting in the highest overall sensitivity at the end of the experiment.

Inspector sensitivity was apparently a function of both KR and task difficulty. When paired together, KR and the High Difficulty task produced higher inspector sensitivity than KR alone. Since the discriminability between defects and nondefects was lower in the High Difficulty task, inspectors committed more errors during task learning. However, for inspectors receiving KR, these errors were now known and could be used to improve subsequent performance. Presumably, each error forced the inspector to reevaluate and adjust his/her mental model used to detect defects. While inspectors trained with KR and the High Difficulty task were expected to have lower sensitivity than those performing the Low Difficulty task during training, this trend was reversed when performing the standard No KR/Low Difficulty task during Phases 2 and 3. Inspectors trained on the Low Difficulty task with KR made less errors during task learning and, therefore, had less opportunities to update and refine their mental model. When transferred to Phase 2 without KR, Low Difficulty inspectors possessed a less refined mental model which lowered sensitivity.

KR utilization appears to partly depend on the level of task difficulty. In this context, task difficulty is assumed to be directly related to inspector "effort" since the more errors, the more updates required to an individual's mental model, and, therefore, the more effort required (assuming equally motivated subjects). Therefore, the ability of KR to produce accurate inspection decisions is mediated, to some extent, by the level of inspector effort used during training.

Inspectors trained without KR had expected and consistent performance trends. Group I inspector's (No KR/Low Difficulty) sensitivity remained constant between both phases and probability conditions. For Group II, inspector performance was expectantly low during training on High Difficulty defects, however, in subsequent phases sensitivity increased to Group I levels and remained stable. Although Group II inspectors also committed more errors during training, without KR, these errors could not be used to improve their mental models and, consequently, sensitivity was reduced during Phases 2 and 3.

These results support the KR utilization model advanced in Chapter 1. Inspector performance as conceptualized by the SDT parameter, d', was significantly and consistently increased in the presence of KR. KR provided critical information on the characteristics and limits of defective line segments (defect specification knowledge) by identifying errors to the inspector and allowing him/her to successfully distinguish defects from nondefects. This knowledge forms the basis for the inspector's mental model for detecting defects. This model may be thought of as a visual image or "template" which is used to compare each inspection item and evaluate the degree of "defectiveness" present. During inspection, KR is used by the inspector to adjust his/her template to correspond as closely as possible to the shortest item that would be reported "defect". Items judged shorter than this template will be reported as nondefects while items judged the same or longer will be reported as defects. For inspectors trained without KR, this template is less developed and more variable from trial to trial. Inspector sensitivity, therefore, would be expectantly lower as verified by these experimental results.

The other major SDT parameter used to assess inspector performance was response bias (ß). As opposed to d', higher performance on ß was associated, not with absolute magnitude, but rather on optimal placement for a given set of conditions. As inspectors were consistently transferred from 0.2 to 0.4 defect probability conditions across the 3 phases, optimal ß (ß*) varied between 4.0 and 1.5. Since at no time did inspectors exactly match ß*, most of the ß analysis relied on trends toward optimality. For example, inspectors trained with KR were more likely to shift their ß's in the optimal direction on the second replication of the trial than NO-KR inspectors. Specifically for KR inspectors, optimal shifts occurred on 75% of the second replications while for NO-KR, optimal shifts occurred on less than 10%. In particular, KR inspectors detecting Low Difficulty defects were considerably more optimal in ß placement during training and shifting ß's during Phases 1 and 2 than those detecting High Difficulty defects (see Figure 16).

While KR appeared to contribute toward more optimal inspection decisions, defect difficulty also had a strong mediating effect. High Difficulty defects produced very

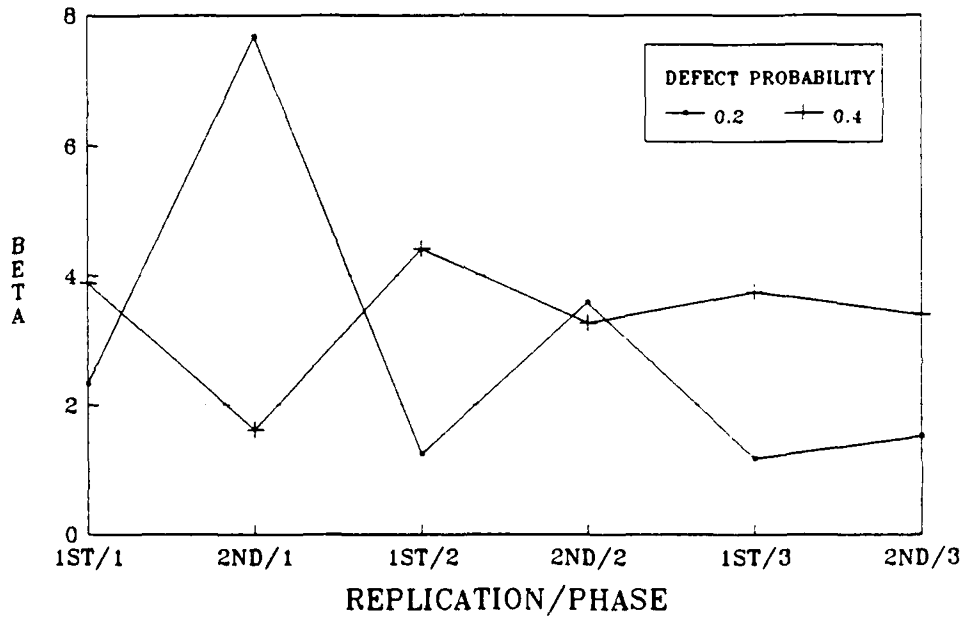


Figure 16. Replication/Phase X Defect Probability for Group III ß

consistent, insensitive ß's during training regardless of the presence of KR or probability condition. During Phases 2 and 3, there was a tendency for ß in the 0.4 probability condition to become relatively higher or more conservative than the 0.2 condition, although this difference was much greater for KR inspectors. This relationship was also apparent and fairly consistent for Low Difficulty defects although, again, this difference was much greater for KR inspectors. The explanation for this result lies in the consistent presentation of defect probability conditions. Inspectors performing 2 replications of 0.2 first become "primed" to the lower defect rate, inhibiting the tendency to report defects in the 0.4 condition and resulting in a higher, more conservative ß. During training on Low Difficulty defects, this effect disappears with KR. This effect is also more pronounced for KR inspectors during Phases 2 and 3. Apparently, inspectors initially trained with KR become more insensitive to changes in defect probability, although training on Low Difficulty defects permits more ß optimal shifts on the second replication. It appears that KR established a very strong bias for the existing probability conditions that results in nonoptimal inspection decisions as defect probabilities change; however, for a given probability condition, KR inspectors are more likely to shift their ß's in a more optimal direction on the second replication than NO-KR inspectors.

Chapter 6

OVERALL DISCUSSION AND CONCLUSIONS

In the course of analyzing the data from the three previous experiments, it was generally assumed that the two major parameters of SDT, d' and ß, were relatively independent measures of inspector performance. This assumption is inherent in the SDT model and was confirmed by the experimental data. As a result, the inspection skill reflected by each of these parameters will be discussed separately within its own model and then integrated within the conclusions section.

Sensitivity

One major result of this series of experiments was the clear and consistent finding that KR increased sensitivity, as measured by d', of visual inspectors detecting line length differences on a computer screen. Overall, KR increased inspector sensitivity by an average of over 23% compared to NO-KR inspectors. This superiority of KR was consistent across both defect difficulty and probability levels, although Experiment 1 results suggested that the increase was more dramatic for Low Difficulty defects and high probability levels.

Experiment 1 results also showed that inspectors had higher sensitivity for low difficulty defects when they were preceded by high difficulty defects than when presented first. Interpreted within attention theory (Lintern and Wickens, 1987), this result provided evidence that task learning was enhanced by "training" on a higher difficulty version of the task, if the source of difficulty directly contributed to task learning. This finding was confirmed in Experiment 3 where inspectors trained with KR and High Difficulty defects attained the highest sensitivity of any training group.

While the superiority of KR was well established in the first experiment, Experiment 2 demonstrated that this advantage could not be explained solely in terms of inspector motivation. At least part of the KR effect was to transmit information to the inspector which increased his/her ability to discriminate defects from nondefects. This information is believed to be derived from the inspector's awareness of errors which is then used to make internal model adjustments and more accurate decisions.

In a transfer of training paradigm used in Experiment 3, inspector sensitivity was increased both during training when KR was present and in later sessions when KR was removed. This superiority of KR was maintained when inspectors were tested immediately after training and three weeks later without KR, especially for inspectors trained on High Difficulty defects. Apparently higher difficulty levels allowed inspectors to more effectively process KR information and enhanced task learning.

The basis of any model for explaining KR effects on sensitivity must address the nature of the inspector's internal representation of task events. The idea that KR increases sensitivity by increasing habit strength through reinforcement (KR) has been discounted and confirmed by the results of Experiment 2. A more information processing approach uses the construct of a "perceptual trace" (Adams, 1987) to represent the internal trial by trial model of an inspector concerning the perceived distinction between defects and nondefects. During advanced stages of learning, the perceptual trace may be stored in long term memory as a "template" which is down loaded at the beginning of an inspection session. The perceptual trace is conceived of as a "working copy" of the template in memory and which can be changed or altered on a trial by trial basis.

The template develops during the initial stages of learning when inspectors are simply observing the line length judgement task. As the inspector practices the task, the template in memory is down loaded and adjusted as a perceptual trace. The most sensitive templates have lengths which are closest to 50% of the length of the viewing area or, in other words, the shortest defective line segment possible. With KR comes knowledge of errors and adjustments of the perceptual trace to correspond more closely to the 50% length. "Defect" response errors shorten the perceptual trace and "nondefect" errors lengthen it.

Manipulating task difficulty by varying the discriminability of defects is a positive influence on sensitivity, especially when paired with KR. Inspectors performing higher difficulty versions of the task were forced to make finer adjustments, enhanced with KR, of their perceptual trace on a trial by trial basis. The resulting template was more sensitive than one obtained by only performing Low Difficulty inspection tasks.

Increasing defect probability had only minor effects on inspector sensitivity. Although many vigilance studies predicted improved observer performance with increasing defect probability (Stroh, 1971), this improvement was usually characterized by an increase in HR without necessarily checking the corresponding FAR. As a result, many of these so-called "improvements" in performance may result in a decrease in ß rather than a real increase in sensitivity (Parasuraman and Davies, 1976; Swets, 1977). The experimental results reported here generally supported the SDT model assumption that d' and ß are relatively independent measures of inspector performance.

Sensitivity results were generally consistent from Phases 2 to 3. In other words, inspector retention of sensitivity skill remained essentially unchanged during the course of the three-week interval between measurements. Such short term stability in performance is not uncommon (Goldberg and O'Rourke, 1989). Skill retention was particularly dramatic for Group IV inspectors (KR/Low Difficulty defects) whose mean d' during Phase 3 was the highest recorded during the experiment. The retention of inspection skill for longer intervals is a topic for future research.

Response Criterion

Response criterion effects in this experiment were mixed and less clear-cut than the sensitivity results. According to the SDT model, an inspector should adjust his response criterion, measured by ß, as defect probabilities and decision payoffs change, while keeping sensitivity relatively constant. Optimally, an inspector should decrease ß as defect probabilities increase and decision payoffs favor more hits, and increase ß as defect probabilities decrease and payoffs favor less false alarms. Prior data, however, showed that human inspectors are conservative decision makers who adjusted their ß's less than that normatively expected (Baddeley and Colquhoun, 1969). There are two primary methods used to assess response criterion performance: First, observing directional changes in ß (increase or decrease) as probabilities or payoffs changed without regard to the specific values of ß taken on; and second, converting ß values to optimality scores computed by $|\beta - \beta^*|$.

Experiment 1 results provided solid evidence that inspectors using KR manipulated ß more optimally, both in terms of directional changes and optimality scores. An unexpected finding was that ß was significantly lower and more optimal for High Difficulty defects. One interpretation of this result is that inspectors faced with higher difficulty tasks experience greater uncertainty about their decision and have a larger tendency to respond "defect" on a given trial. This tendency is probably due to an inspector's prior expectancy of defects which is high in the absence of contradictory information. Since inspectors are not obtaining sufficient probability information from the more difficult tasks to shift their ß more optimally, they are forced to rely on inaccurate and insufficient prior expectancies (which are usually more liberal) to make their decisions. The observed result of greater optimality for High Difficulty defects may be due to levels of defect probabilities selected rather than any real main effect. Inspectors with overall lower ß's, regardless of their knowledge of defect probabilities, will have overall more optimal performance in spite of being virtually insensitive to any probability change.

The sequence of difficulty levels disrupted ß adjustments. When the low difficulty task was presented first, inspectors receiving KR decreased their ß's as defect probabilities increased in accordance with the SDT model but not as much as normatively predicted. When the High Difficulty task was presented first, ß adjustments were completely inaccurate with ß's increasing significantly for NO-KR inspectors and remaining constant with KR as defect probabilities increased. Apparently, the sensitivity advantage of Low to High Discriminability Sequence (High to Low difficulty) does not carry over to response criterion performance. Extending the attention theory explanation, inspecting High Difficulty defects, while providing inspectors with more opportunities to learn the more subtle differences between defects and nondefects, failed to provide a similar advantage for response criterion performance. This is consistent with the SDT model which assumes that the parameters d' and ß, and the skills which underlie their changes, are relatively independent.

Experiment 2 results reinforce the superiority of KR on response criterion performance. Eoth KR groups had overall more optimal ß adjustments than NO-KR inspectors. However, when all six optimal ß's are plotted, one for each payoff/defect probability combination, only TRUE-KR inspectors adjusted their ß's such that all six were not significantly different from ß*. Varying decision payoffs was more effective and relevant to inspectors for optimally adjusting ß than defect probabilities alone.

Experiment 2 results also showed that NO-KR inspectors had significantly higher ß in the 0.4 probability condition. Initially, inspectors set ß relatively low anticipating a high number of defects. This prior expectancy is probably due to the nature of the task, which was to detect defects. Without KR, inspectors were unable to obtain current probability information to optimally adjust ß for the given probability condition. Since inspectors were always presented with probability sequence 0.2 then 0.4, ß tended to be much lower than predicted by the ß* model for the 0.2 probability condition. As the number of defects presented doubled in the 0.4 condition, inspectors probably became aware that many of the previous defects reported under the very liberal initial criteria were actually nondefects. The result was an attempt to correct perceived errors by a general inhibition of the defect response and a higher (more conservative) ß. Higher ß was also observed in the 0.4 condition during Experiment 1 for NO-KR inspectors in the High to Low difficulty sequence.

Experiment 3 results tracked inspector ß performance through four different training groups, which varied the presence of KR and difficulty level across the three phases. The effects of training could then be evaluated immediately after training and three weeks later on a standard NO-KR/Low Difficulty task. Across all phases, inspectors trained without KR had generally lower ß's than KR trained inspectors, especially for High Difficulty defects. From the previous discussion, inspectors tend to lower ß when uncertain about a decision and emphasize detecting defects (hits). While ß increased overall as defect probabilities increased from 0.2 to 0.4, this trend reversed during the

Training phase as a result of KR/Low Difficulty trained inspectors. The superiority of KR coupled with Low Difficulty defects for producing more optimal ß adjustments was again demonstrated.

The effect of inspector training on optimal ß adjustment was not significant, although performance was more optimal for the 0.4 defect probability condition, especially for NO-KR trained inspectors. Again, a situation similar to Experiment 1 exists whereby inspectors experiencing greater uncertainty (with NO-KR) about the current probability conditions resort to an overall lower (more liberal) ß which happens to be closer to ß*, but which is insensitive to changing probability conditions.

Since Experiment 3 included 2 replications for each experimental condition, it was possible to more closely examine ß adjustments. The distinction between Local Probability (LP) knowledge and Cumulative Probability (CU) knowledge was first made by Vickers, Learly, and Barnes (1977) in criticizing the ideal observer hypothesis (Williges, 1976). LP knowledge represents the trial by trial knowledge of defect probabilities while CP knowledge represents knowledge of defect probabilities obtained from the very beginning of the experimental session including prior expectancies and training.

During the training phase, which was the only time when KR and task difficulty manipulations were present, Group III inspectors (KR/Low Difficulty defects) shifted ß in the optimal direction both within and between probability conditions. The data showed that ß increased between replications of the 0.2 probability condition and decreased in both replications of the 0.4 condition. The remaining training groups had relatively low and insensitive ß's. In the Phase 2 immediate post-test (NO-KR/Low Difficulty defects), Group III inspectors continued to maintain optimal performance trends, while the ß's for the NO-KR Groups (I and II) remained low and generally insensitive. Group IV inspectors, however, experienced a dramatic rise in ß from the end of the last 0.2 replication to the end of the first 0.4 replication. During Phase 3, three weeks later, Group IV continued this upward trend across conditions. Group III inspectors also displayed a similar increase in ß between probability conditions, and both experienced a general flattening of their ß's within each probability condition. Both Groups I and II continued to have very low and insensitive ß's across all probability conditions.

Withdrawing KR apparently caused inspectors to adopt very conservative (high ß) criterion as defect probability increased from 0.2 to 0.4, especially when trained initially on High Difficulty defects. Since KR is believed to be used by an inspector to update his LP knowledge, when it is no longer available, an inspector must rely on his perceived CP knowledge to make a decision. This knowledge is more accurate for Group III inspectors trained with KR on Low Difficulty defects. When KR is removed, the superior CP knowledge allows more optimal decision making as shown in these results. However, as time goes by, the memory trace deteriorates and inspectors again move to more conservative decision making as defect probabilities increase. Group IV inspectors performing the High Difficulty task with KR, had the ability to update LP knowledge but lacked the ability to use it in the High Difficulty condition. Consequently, β remained fairly constant during Phase 1. While performing the NO-KR/Low Difficulty task during Phase 2, Group IV inspectors initially adopted a liberal criterion without KR for the 0.2 condition. However, by the end of the first replication of the 0.4 condition, inspectors now performing the Low Difficulty task realized that many previously identified "defects" were false alarms and in an attempt to correct for these errors, inspectors increased β to compensate. The result was a general tendency to inhibit the "defect" response and increase ß are defect probabilities increased. This trend remained consistent during Phase 3, three weeks later.

In general, response criterion results remained unchanged from Phase 2 to Phase 3. Inspectors tested immediately after training and again three weeks later showed little change in response criterion performance. In fact, any small changes observed were usually in the positive direction. However, deterioration of inspection skill over longer retention intervals may be more important.

Inspector Latency

The results of this experiment showed inconsistent effects on RT. In Experiment 1, KR clearly resulted in faster RT's for both task difficulty levels, although the effect was more dramatic for High Difficulty defects. On the other hand, RT's were significantly slower for High Difficulty defects when presented first compared to after Low Difficulty defects. RT's were overall slower for High Difficulty defects.

Inspector RT was not significantly effected by any of the experimental variables in Experiment 2. Mean RT was faster for both KR groups compared to NO-KR, although the difference was not significant. The KR X Defect Probability interaction approached significance suggesting that KR lowered RT more for 0.2 compared to 0.4 defect probability conditions.

In Experiment 3, KR trained inspectors had overall faster RT's than NO-KR, especially for Low Difficulty defects. RT's were also significantly faster in Phase 3 and for 0.4 defect probability. This reduction in RT as defect probabilities increased was only observed for NO-KR inspectors. As inspectors progressed through the phases, NO-KR inspectors tended to decrease RT's while KR inspectors tended to increase, although the changes were small and nonsignificant.

The most consistent finding in the above experiments is that KR generally reduced the time to make an inspection decision. The superior template developed from KR can be used to make faster as well as more accurate inspection decisions. No-KR inspectors are forced to compare line segments with either external reference points or very imprecise internal representations. In either case, more time is taken on average to make a decision compared to the time needed to make a single comparison against one very accurate defect template.

The effect of task difficulty on RT is inconclusive. For example, Group I inspectors in Experiment 3 had significantly slower RT's than Group II even though Group II inspected higher difficulty defects. It seems that RT's under some circumstances may be effected by inspector arousal which can be assumed to be lower for Low Difficulty defects and lower defect probability levels. However, increased sensitivity with KR is generally accompanied by faster RT's.

Conclusions

Based on results of the 3 experiments discussed above, a clearer understanding of KR utilization in visual inspection has been obtained. Ten broad conclusions can be integrated within a general model of inspector performance:

1. KR improved inspector sensitivity (d') by making inspectors aware of errors which forced template adjustments and resulted in more accurate decisions.

2. KR resulted in faster inspector decisions as the superior template reduced the number of individual operations needed, and hence, the time necessary to make a decision.

3. KR provided local probability information used to shift ß in the more optimal direction as defect probabilities increased.

4. Removal of KR maintained the sensitivity effect while criterion placement was disrupted for High Difficulty defects.

5. High task difficulty, as manipulated by defect discriminability, resulted in higher sensitivity with KR but less optimal ß shifts as probabilities increased.

6. Manipulating decision payoffs was more effective in optimally adjusting ß than defect probabilities.

7. No-KR inspectors tended to increase ß as defect probabilities increased from 0.2 to 0.4.

8. Defect probabilities had minimal effect on inspector sensitivity.

9. Inspector sensitivity (d') and response criterion (ß) were relatively independent measures of inspection performance.

10. Inspection skill retention did not deteriorate after 3 weeks.

Based on the above conclusions, a model of KR utilization is proposed which distinguishes sensitivity from response criterion knowledge and local probability from cumulative probability knowledge. Figure 17 shows one idea of how these constructs can be combined and structured to explain the experimental results obtained. Inspection items are perceived by a human inspector and the sensory information (visual, in this case) flows into working memory via some type of processor. For the line segment detection task used here, activated areas of long term memory, which include information germane to the task such as perceived prior probabilities and visual templates stored, are also down-loaded to working memory. The visual information obtained from an inspection item is compared to the template stored in memory to determine if the line segment is long enough to be called a defect. In addition, perceived probabilities are also being considered before the actual decision is made. If the sensory information is compelling, the inspector will probably base his decision primarily on this information alone. If the information is uncertain, perceived probabilities will become more important. Inspectors generally base their decision tendencies on their cumulative probability knowledge unless local probability knowledge is available, usually through KR.

When No-KR or High Difficulty defects are present, inspectors are unable to adequately develop or use local probability knowledge. As a result, inspection decisions

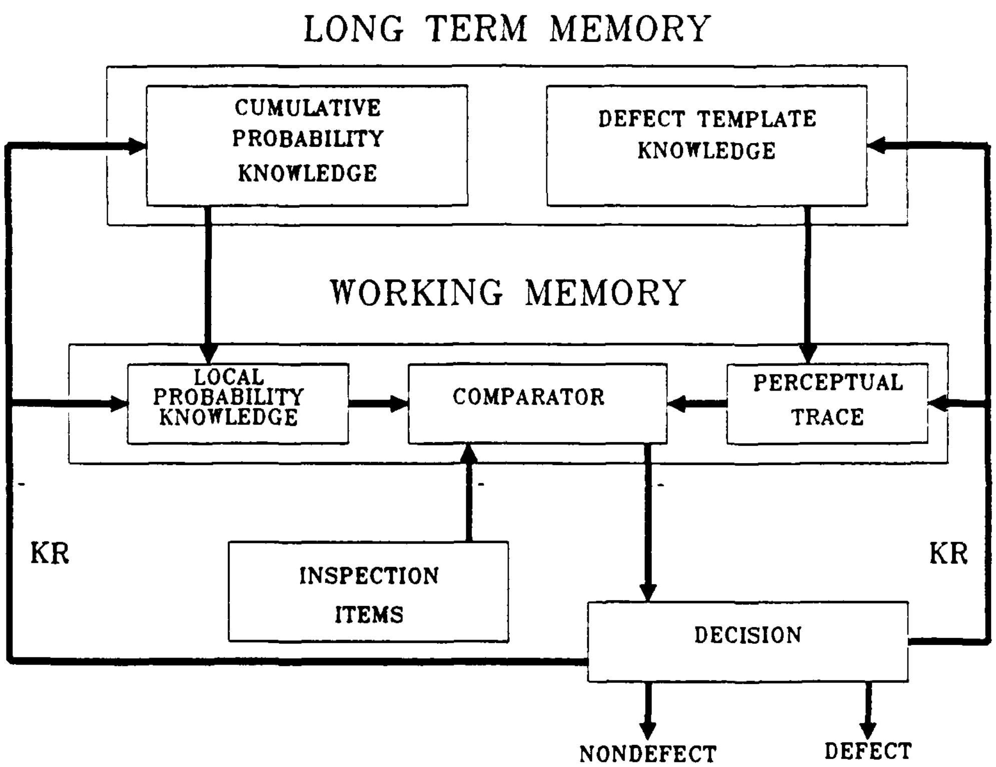


Figure 17. Revised Decision Making Model for Visual Inspection

are based primarily on cumulative probability knowledge which consists of an amalgamation of previous experiences (e.g., prior expectations and training stored in memory.

Most inspectors initially set ß low at the outset of the experiment due to the prior expectation that detection of defects is more important than false alarms. If the inspector received KR, ß quickly recovered to reflect the ongoing defect rate and also adapted to increases in defect probability, for Low Difficulty defects. For High Difficulty defects, LP knowledge was disrupted due to greater attention given to sensitivity performance and ß remained low and insensitive to defect probability changes. Without KR, ß increased as defect probabilities increase, in conflict with the ideal observer hypothesis. As the number of defects increased in the 0.4 probability condition, inspectors corrected for their initially overly liberal ß by inhibiting the "defect" response and increasing ß.

In addition, once KR is removed, inspectors trained on High Difficulty defects tend to become more conservative decision makers as defect probabilities increase. Since defect probabilities in this experiment were always presented as increasing from 0.2 to 0.4, inspectors may be primed by the previously low defect rate to respond more conservatively even when the number of defects increased.

Having been trained with KR, these inspectors are now deprived of the only information source they had to make decisions. When training also involved Low Difficulty defects, inspectors were able to more efficiently learn to use LP knowledge and shift ß more optimally even when KR was removed (although this ability seemed to degrade somewhat during Phase 3). However, when training also involved High Difficulty defects, inspectors' abilities to develop LP knowledge and accurately shift ß during training were severely degraded. When KR was removed and defects became easier, these inspectors tried to replace the information they were lacking by drawing on CP knowledge which contained mostly low probability information from the first two 0.2 trials. When defect probability increased to 0.4, inspectors ignored any LP knowledge available from the easier defects and relied completely on lower probability contents of CP knowledge, resulting in higher ß.

Thus, KR provided information used by the inspector to not only manipulate ß optimally based on the ongoing defect rate, but to also shift ß more optimally as defect probabilities increased. However, increasing task difficulty may negate this advantage. The interaction between task difficulty and KR is complex and further research is necessary.

REFERENCES

Adams, J.A. (1987). Historical review and appraisal of research on the learning, retention, and transfer of human motor skills. Psychological Bulletin, 101(1), 41-74.

Adams, J.A., and Humes, J.M. (1963). Monitoring of complex visual displays IV: Training for vigilance. Human Factors, 5(1), 147-153.

Adams, S.K. (1975). Decision making in quality control: Some perceptual and behavioral considerations. In C.G. Drury and J.G. Fox (Eds.), Human reliability in quality control (pp. 55-70). London: Taylor and Francis.

Annett, J. (1966). Training for perceptual skills. Ergonomics, 9, 459-468.

Antonelli, D.C., and Karas, G.G. (1967). Performance on a vigilance task under conditions of true and false knowledge of results. Perceptual and Motor Skills, 25, 129-138.

Baker, E.M. (1975). Signal detection theory analysis of quality control inspector performance. Journal of Quality Technology, 7, 62-71.

Baddeley, A.D., and Colquohoun, W.P. (1969). Signal probability and vigilance: A reappraisal of the 'signal rate' effect. British Journal of Psychology, 60, 165-178.

Bisseret, A. (1981). Application of signal detection theory to decision making in supervisory control. Ergonomics, 24, 81-94.

Buck, L. (1968). Reaction time as a measure of perceptual vigilance. Psychological Bulletin, 65, 291-308.

Colquhoun, W.P. (1975). Evaluation of auditory, visual and dual mode displays for prolonged sonar monitoring in repeated sessions. Human Factors, 17, 425-437.

Czaja, S.J., and Drury, C.G. (1981). Training programs for inspection. Human Factors, 23, 473-484.

Davies, D.R., and Parasuraman, R. (1982). The psychology of vigilance. London: Academic Press.

Davies, D.R., and Tune, G.S. (1970). Human vigilance performance. London: Staples.

Drury, C.G. (1975). Human decision making in quality control. In C.G. Drury and J.G. Fox (Eds.), Human reliability in quality control (pp. 45-54). London: Taylor and Francis.

Drury, C.G. (1982). Improving inspection performance. In Salvendy, G. (Ed.), Handbook of industrial engineering. New York: Wiley.

Drury, C.G. (1988). Personal communication.

Drury, C.G., and Addison, S.L. (1973). An industrial study of the effects of feedback and fault density on inspection performance. Ergonomics, 16, 159-169.

Drury, C.G., and Fox, J.G. (1975). Human reliability in quality control. London: Taylor and Francis.

Drury, C.G., and Sinclair, M.A. (1983). Human and machine performance in an inspection task. Human Factors, 25, 391-399.

Edwards, W. (1982). Conservatism in human information processing. In D. Kahneman, P. Slovic, and A. Tversky (Eds.), Judgement under uncertainty: Heuristics and biases (pp. 359-369). Cambridge, U.K.: Cambridge University Press.

Embrey, D.E. (1975). Training the inspector's sensitivity and response bias. In C.G. Drury and J.G. Fox (Eds.), Human reliability in quality control (pp. 123-131). London: Taylor and Francis.

Embrey, D.E. (1979). Approaches to training for industrial inspection. Applied Ergonomics, 10, 139-144.

Fleishman, E.A., and Parker, J.F., Jr. (1962). Factors in the retention and relearning of perceptual-motor skill. Journal of Experimental Psychology, 64(3), 215-226.

Fox, J.G. (1975). In C.G. Drury & J.G. Fox (Eds.), Human reliability in quality control (pp. 89-90). London: Taylor and Francis.

Gentner, D., and Stevens, A.L. (Eds.). (1983). Mental models, Hillsdale, NJ: Erlbaum.

Goldberg, J.H., and O'Rourke, S.A. (1989). Prediction of skill retention and retraining from initial training. Perceptual and Motor Skills, 69, 535-546.

Green, D.M. (1960). Psychoacoustics and detection theory. Journal of the Acoustical Society of America, 32, 1189-1203.

Green, D.M., and Swets, J.A. (1966). Signal detection theory and psychophysics. New York: Wiley.

Grindley, G.C., and Townsend, V. (1970). Visual search without eye movement. Quarterly Journal of Experimental Psychology, 22, 62-67.

Hagman, J.D., and Rose, A.M. (1983). Retention of military tasks: A review. Human Factors, 25(2), 199-233.

Harris, D.H. (1966). Effect of equipment complexity on inspection performance. Journal of Applied Psychology, 50, 236-237.

Harris, D.H., and Chaney, F.B. (1969). Human factors in quality assurance. New York: Wiley.

Kahneman, D. (1973). Attention and effort. Englewood Cliffs, N.J.: Prentice Hall.

Kahneman, D., Slovic, P., and Tversky, A. (1986). Judgement under uncertainty: Heuristics and biases. Cambridge, U.K.: Cambridge University Press.

Kelley, C.R. (1969). What is adaptive training? Human Factors, 11, 547-556.

Kieras, D.E., and Bovair, S. (1984). The role of a mental model in learning to operate a device. Cognitive Science, 8, 255-273.

Lachman, R., Lachman, J.L., and Butterfield, E.C. (1979). Cognitive psychology and information processing. New York: Erlbaum.

Lintern, G., and Wickens, C.D. (1987). Attention theory as a basis for training research. University of Illinois: Aviation research laboratory technical report ARL-87-2/NASA-87-3.

Mackworth, J.F. (1964). The effect of true and false knowledge of results on the detectability of signals in a vigilance task. Canadian Journal of Psychology, 18, 106-117.

Mane, A., and Wickens, C.D. (1986). The effects of task difficulty and workload on training. In Proceedings of the Human Factors Society 30th Annual Meeting (pp. 1124-1127). Santa Monica, CA: Human Factors Society.

Martin Marietta (1981). Workmanship standards, electrical and mechanical. Orlando Aerospace: Orlando, FL.

Moll, R.A. (1976). Product liability: A look at the law. Engineering Education, 66, 326-331.

Montgomery, D.C. (1984). Design and Analysis of Experiments. New York: Wiley.

Moray, N. (1967). Where is capacity limited? A survey and a model. Acta Psychologica, 27, 84-92.

Moray, N. (1981). The role of attention in the detection of errors and the diagnosis of failures in man-machine systems. In J. Rasmussen & B. Rouse (Eds.), Human detection and diagnosis of system failures. New York: Plenum Press.

Naylor, J.C., Briggs, G.E., and Reed, W.G. (1968). Task coherence, training time, and retention interval effects on skill retention. Journal of Applied Psychology, 52(5), 386-393.

Neter, J., Wasserman, W., and Kutner, M.H. (1985). Applied linear statistical models. Homewood, IL: Irwin.

Ono, H. (1967). Difference threshold for stimulus length under simultaneous and nonsimultaneous viewing conditions. Perception and Psychophysics, 2, 201-207.

Parasuraman, R. and Davies, D.R. (1976). Decision theory analysis of response latencies in vigilance. Journal of Experimental Psychology: Human Perception and Performance, 2, 578-590.

Peterson, C.R., and Beach, L.R. (1967). Man as an intuitive statistician. Psychological Bulletin, 68, 29-46.

Poulton, E.C. (1973). Unwanted range effects from using within subject experimental designs. Psychological Bulletin, 80, 113-121.

Roehrig, W.C. (1964). Psychomotor task with perfect recall after fifty weeks of no practice. Perception and Motor Skills, 19, 547-550.

Rouse, W.B., and Hunt, W.B. (1981). Problem solving skills of maintenance trainees in diagnosing faults in simulated power plants. Human Factors, 21, 195-204.

Schneider, W. (1985). Training high performance skills: Fallacies and guidelines. Human Factors, 27(3), 285-300.

Smith, R.L., and Lucaccini, L.F. (1969). Vigilance research: Its relevance to industrial problems. Human Factors, 11, 149-156.

Sternberg, S. (1967). Two operations in character recognition: Some evidence from reaction time measurements. Perception and Psychophysics, 2, 45-53.

Sternberg, S. (1969). Memory scanning: Mental processes revealed by reaction time measurements. American Scientist, 57, 421-457.

Stroh, C.M. (1971). Vigilance: The problem of sustained attention. New York: Pergamon Press.

Swets, J.A. (Ed.). (1964). Signal detection theory and recognition by human observers: Contemporary readings. New York: Wiley.

Swets, J.A. (1966). Deferred decision in signal detection. In B.F. Lomov (Ed.), Detection and recognition of signals (pp. 21-26). Lenin Hills, Moscow: Moscow University Press.

Swets, J.A. (1977). Signal detection theory applied to vigilance. In R.R. Mackie (Ed.), Vigilance: Theory, Operational Performance, and Physiological Correlates (pp. 705-718). New York: Plenum.

Swets, J.A., Tanner, W.P., and Birdsall, T.G. (1961). Decision processes in perception. Psychological Review, 78, 301-340.

Thresh, J.L., and Frerichs, J.S. (1966). Results through management application of human factors. Paper presented at the American Society of Quality Control Technical Conference, New York.

Tversky, A., and Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. Science, 185, 1124-1131.

Van Cott, H.P., and Kinkade, R.G. (1972). Human engineering guide to equipment design. Washington, D.C.: U.S Government Printing Office.

Veldhuyzen, W., and Stassen, H.G. (1976). The internal model: What does it mean in human control? In T.B. Sheridan & G. Johannsen (Eds.), Monitoring behavior and supervisory control. New York: Plenum Press.

Vickers, D., Learly, J., and Barnes, P. (1977). Adaptation to decreasing signal probability. In R.R. Mackie (Ed.), Vigilance: Theory, Operational Performance, and Phsyiological Correlates (pp. 679-704). New York: Plenum.

Wallach, P.M., and Adams, S.K. (1969). The utility of signal detection theory in the analysis of industrial inspection accuracy. AIIE Transactions, 1, 33-44.

Wang, M.J., and Drury, C.G. (1987). A method of evaluating job demands in industrial inspection. In S.S. Asfour (Ed.), Trends in ergonomics/human factors IV (pp. 89-96). New York: Elsevier Science.

Warm, J.S., Epps, B.D., and Ferguson, R.F. (1974). Effects of knowledge of results and signal regularity on vigilance performance. Bulletin of the Psychonomic Society, 4, 272-274.

Weidenfeller, E.W., Baker, R.A., and Ware, J.R. (1962). Effects of knowledge of results (true and false) on vigilance performance. Perceptual and Motor Skills, 14, 211-215.

Welford, A.T. (1968). Fundamentals of skill. London: Methuen.

Wickens, C.D. (1980). The structure of attentional resources. In R. Nickerson & R. Pew (Eds.), Attention and performance VIII. Englewood Cliff, N.J.: Erlbaum.

Wickens, C.D. (1984). Engineering psychology and human performance. Columbus, OH: Charles E. Merrill.

Wickens, C.D., and Kessel, C.J. (1979). The effects of participatory mode and task workload on the detection of dynamic system failures. IEEE transactions on Systems, Man, and Cybernetics, SMC-13, 24-31.

Wiener, E.L. (1968). Training for vigilance: Repeated sessions with knowledge of results. Ergonomics, 11, 547-556.

Williams, P.S. (1986). Processing demands, training, and the vigilance decrement. Human Factors, 28, 567-579.

Williges, R.C. (1973). Manipulating the response criterion in visual monitoring. Human Factors, 15, 179-185.

Williges, R.C. (1976). The vigilance increment: An ideal observer hypothesis. Paper Presented at the NATO Symposium on Monitoring Behavior and Supervisory Control, Berchtesgaden, West Germany.

Appendix A

STATISTICAL MODEL

The ANOVA model used in these experiments was a "mixed" model with both between-subject (nested) and within-subject (blocked) variables (Neter, Waserman, and Kutner, 1985; p. 1021). KR groups were always treated as between-subject variables to avoid the carry-over effects inherent in going from KR to NO-KR. Within-subject levels were counterbalanced where appropriate. Defect probability levels were not counterbalanced since a stated objective was to examine inspector performance as quality deteriorates and defect probability increases.

Experiment 1

This was a four-factor experiment with Factors A and B (KR, Difficulty Level Sequence, respectively) treated as between-subject factors and Factors C and D (Task Difficulty, Defect Probability, respectively) within-subject and completely crossed. Five subjects were assigned to each of four groups with no other replications.

Assuming fixed treatment, random subject effects, and no treatment X subject interactions, the appropriate model is:

$$Y_{ijklm} = U_{....} + A_i + B_j + C_k + D_l + (AB)_{ij} + (AC)_{ik} + (BC)_{jk} + (BD)_{jl} + (CD)_{kl} + (ABC)_{ijk} + (ABD)_{ijl} + (BCD)_{jkl} + (ACD)_{ikl} + (ABCD)_{ijkl} + p_{m(ij)} + e_{(ijklm)}$$

where:

$U_{....}$ = overall constant

A_i = constant such that $\Sigma A_i = 0$

B_j = constant such that $\Sigma B_j = 0$

C_k = constant such that $\Sigma C_k = 0$

D_l = constant such that $\Sigma D_l = 0$

All interaction terms: $[(AB)_{ij}, (AC)_{ik}, \ldots]$ also represent constants subject to the restriction that the sum of all terms over each level of variables included $[\Sigma(AB)_{ij}$ over i and $\Sigma(AB)_{ij}$ over j, . . $] = 0$.

$p_{m(ij)} + e_{(ijklm)}$ are independent and $N(0, \sigma_p)$ and $N(0, \sigma)$ respectively.

Experiment 2

This was a three-factor study with Factor A (KR) between subject and Factors B and C (Payoff and Defect Probability, respectively) within subject and completely crossed with the sequence of Payoff conditions counterbalanced across subjects. Six subjects were randomly assigned to each of three groups with no replications.

Assuming fixed treatment, random subject effects, and no treatment X subject interactions, the appropriate model is:

$$Y_{ijkl} = U.... + A_i + B_j + C_k + (AB)_{ij} + (AC)_{ik} + (BC)_{jk} + (ABC)_{ijk} + p_{l(i)} + e_{(ijkl)}$$

where:

$U...$ = overall constant

A_i = constant such that $\Sigma A_i = 0$

B_j = constant such that $\Sigma B_j = 0$

C_k = constant such that $\Sigma C_k = 0$

All interaction terms: $[(AB)_{ij}, (AC)_{ik}, (BC)_{jk}, (ABC)_{ijk}]$ also represent constants subject to the restriction that the sum of all terms over each level of variables included $[\Sigma(AB)_{ij}$ over i and $\Sigma(AB)_{ij}$ over $j, . . .] = 0$.
$p_{l(i)} + e_{(ijkl)}$ are independent and $N(0, \sigma_p)$ and $N(0, \sigma)$ respectively.

Experiment 3

This was a three factor study with Factor A (Training Group) between subject and Factors B and C (Phase and Defect Probability, respectively) within subject. Five subjects were randomly assigned to each of four groups with two replications in each condition.

Assuming fixed treatment, random subject effects, and no treatment X subject interactions, the appropriate model is:

$$Y_{ijklm} = U_{....} + A_i + B_j + C_k + (AB)_{ij} + (AC)_{ik} + (BC)_{jk} + (ABC)_{ijk} + p_{l(i)} + e_{(ijklm)}$$

where:

$U_{...}$ = overall constant

A_i = constant such that $\Sigma A_i = 0$

B_j = constant such that $\Sigma B_j = 0$

C_k = constant such that $\Sigma C_k = 0$

All interaction terms: [$(AB)_{ij}$, $(AC)_{ik}$, $(BC)_{jk}$, $(ABC)_{ijk}$] also represent constants subject to the restriction that the sum of all terms over each level of variables included [$\Sigma(AB)_{ij}$ over i and $\Sigma(AB)_{ij}$ over j, . .] = 0.

$p_{l(i)}$ + $e_{(ijklm)}$ are independent and N(0, σ_p) and N(0, σ), respectively.

Appendix B

ANOVA AND SDT MODEL ASSUMPTIONS

ANOVA Assumptions

The ANOVA assumptions of normality and equality of variance for the error terms was confirmed using a residual analysis (Montgomery, 1984; p. 85). If residuals are plotted against fitted values, the result should be a structureless plot of residuals about the 0 axis. This indicates constant variance in error terms. To check the normality assumption, residuals are plotted on a rectangular coordinate system against their corresponding z scores. If the normality assumption holds true, the points should fall roughly along a straight line.

Since both hit rate and false alarm rate were the basic measure from which all other measures were derived, they were selected as the dependent measures for checking the ANOVA model assumptions.

Figures 18 and 19 show the residual plots against fitted values for Experiments 1 through 3 for hit rate and false alarm rate respectively.

Figures 20 and 21 show the residual plot against normalized residuals for Experiments 1 through 3 for hit rate and false alarm rate respectively. An examination of these plots shows that the ANOVA assumptions for hit rate and false alarm rate are satisfied. The residual versus fitted values plots were structureless for all three experiments (R^2 values from regression = 0.0%) with little evidence of a particular pattern. All plots of residuals versus normalized residuals fell along a straight line confirming the normality assumption. For HR, the R^2 values for Experiments 1 - 3 were 98.7%, 97.0%, and 98.0%, respectively. For FAR, these values were 99.5%, 95.8%, and 96.8%. The hypothesis of normality was accepted since the corresponding correlation coefficients exceeded the critical value based on a test for normality described in Minitab, (1988; p. 63).

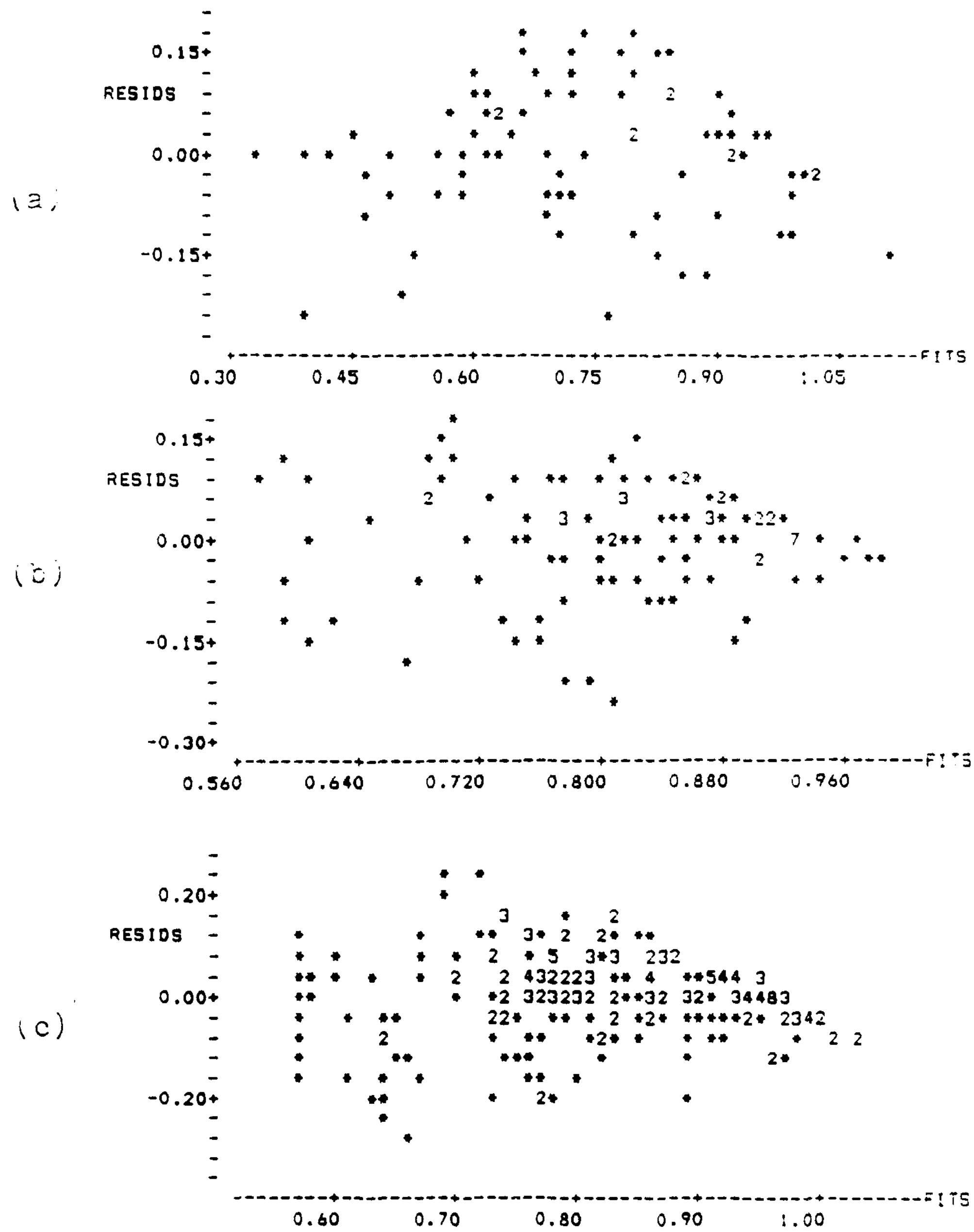


Figure 18. Plots of Residuals By Fitted Values for Hit Rate. (a) Experiment 1 (b) Experiment 2 (c) Experiment 3

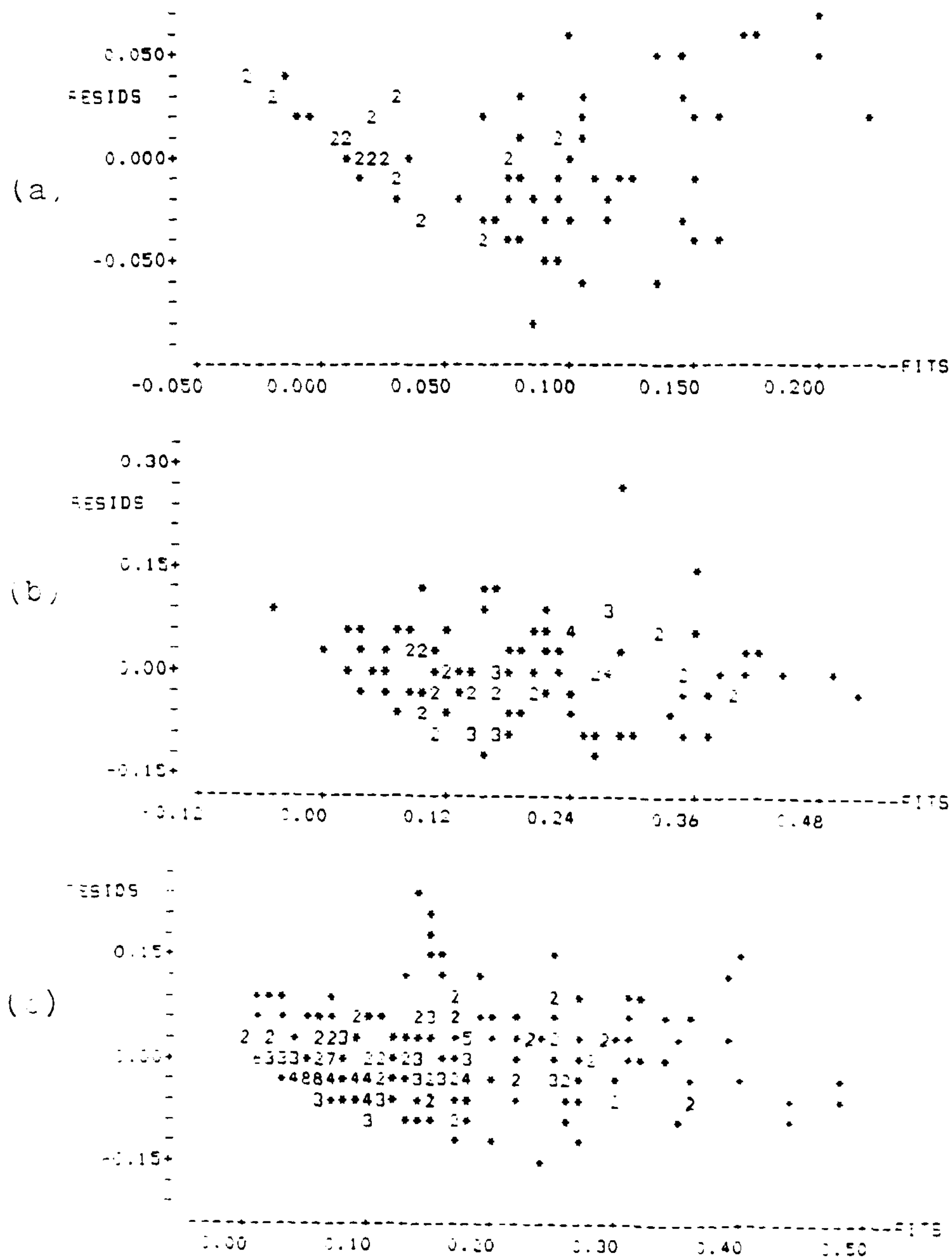


Figure 19. Plots of Residuals By Fitted Values for False Alarm Rate. (a) Experiment 1 (b) Experiment 2 (c) Experiment 3

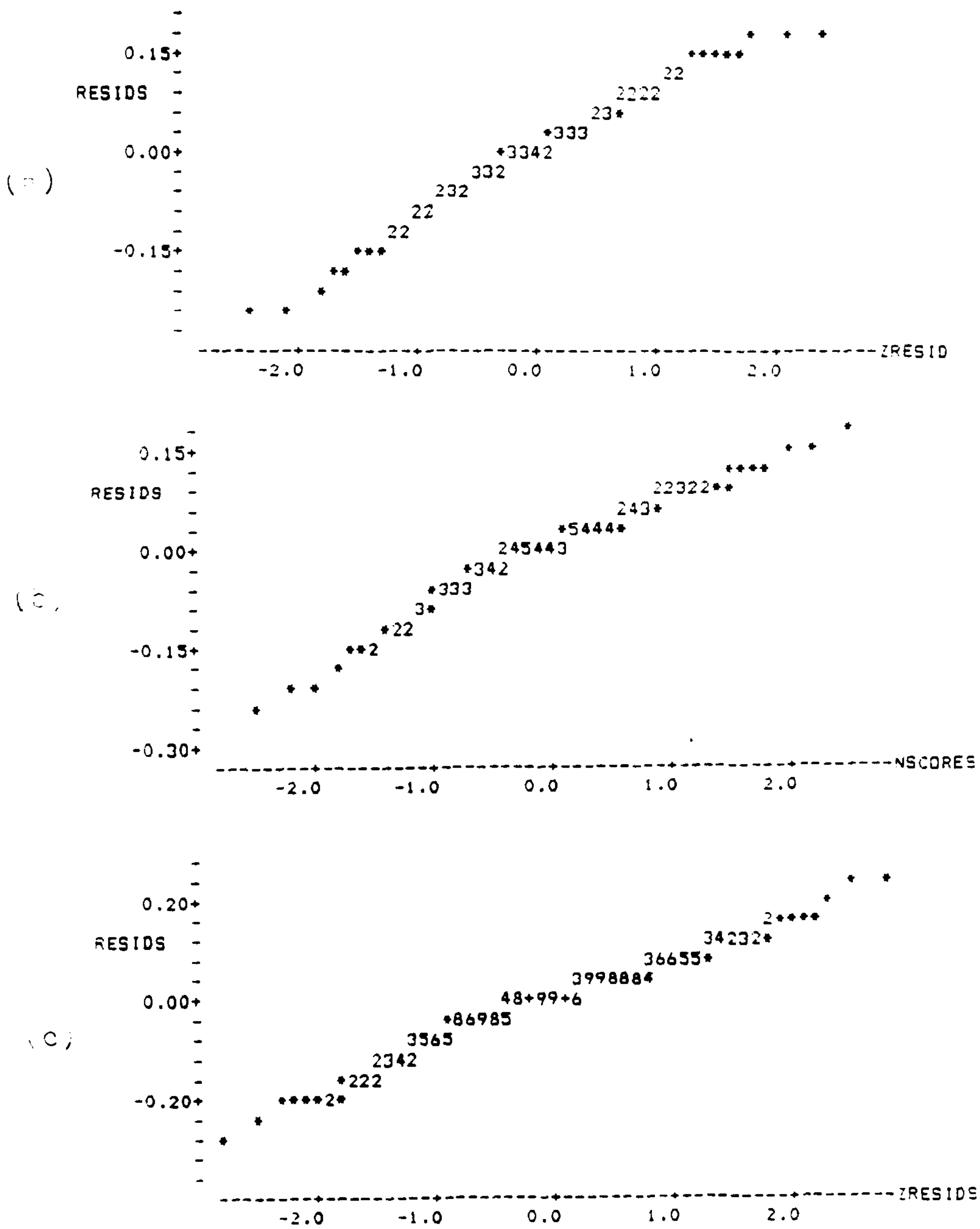


Figure 20. Plots of Residuals By Normalized Residuals for Hit Rate. (a) Experiment 1, R^2 = 98.7% (b) Experiment 2, R^2 = 97.0% (c) Experiment 3, R^2 = 98.0%

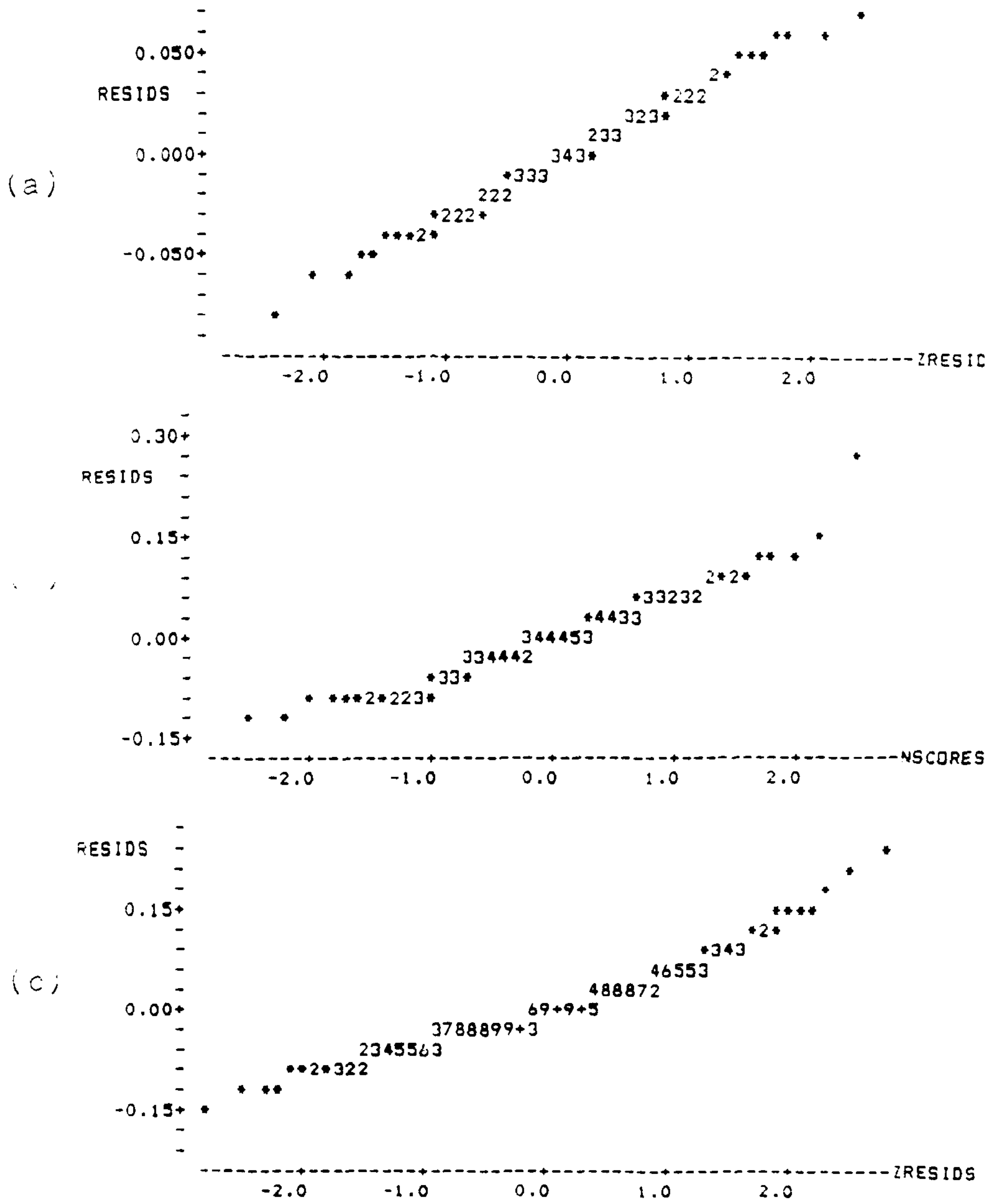
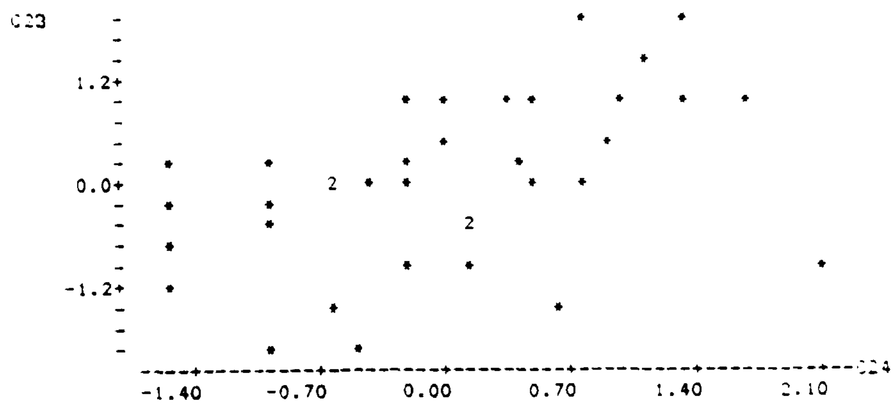


Figure 21. Plots of Residuals By Normalized Residuals for False Alarm Rate. (a) Experiment 1, R^2 = 99.5% (b) Experiment 2, R^2 = 95.8% (c) Experiment 3, R^2 = 96.8%

SDT Assumptions

SDT has significantly contributed to our knowledge of inspection performance by providing a framework for obtaining a measure of inspector sensitivity that is free from the contaminating effects of response bias. Within this framework, d' remains constant as a measure of inspector sensitivity for a given defect, while the inspector's response criterion as measured by ß is free to fluctuate based on the probabilities and payoffs of various decision outcomes. However, this result is only true if the both the nondefect and defect distributions are normal and have equal variance. Baker (1975) recommended converting hit rate and false alarm rate to z scores and plotting these values on rectangular coordinates. If the normality assumption is met, the data points should fall along a straight line. In addition, if the slope of the line equals 1 then changes in FAR z scores produce equal changes in HR z scores and the two distributions can be assumed to have equal variance.

To check the model assumptions, z scores for HR and FAR are plotted for a given level of inspector sensitivity. In Experiment 2, for example, each of the 3 KR groups maintained a constant d' as both payoffs and defect probabilities changed. Within a particular group, it is possible to check the how changes in HR z scores compare to changes in FAR z scores. Figures 22-24 show the plots of these z scores for each group. The scatterplots show fitted lines of positive slope, based on the regression analysis, for all 3 groups. The slopes of these lines, however, were less than 1. This means that the two distributions did not have the same variance. In particular, the variance of the defect distribution was higher than the variance of the nondefect distribution. This is not surprising since the defect distribution was based on a smaller number of observations, hence the greater variance. However, this violation was not large enough to interfere with the assumed independence of d' and ß. Therefore, interpretation of the data within the SDT framework is possible, but caution is required.



The regression equation is
C23 = - 0.008 + 0.471 C24

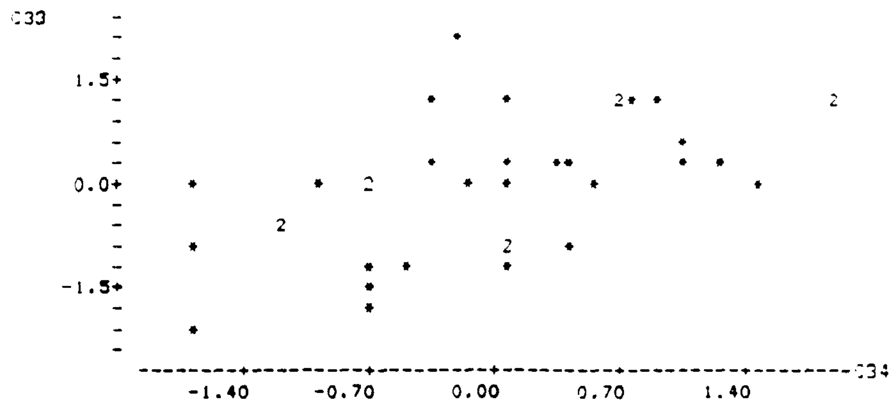
Predictor	Coef	Stdev	t-ratio	p
Constant	-0.0078	0.1422	-0.05	0.957
C24	0.4708	0.1513	3.11	0.004

s = 0.8534 R-sq = 22.2% R-sq(adj) = 19.9%

Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	1	7.0505	7.0505	9.68	0.004
Error	34	24.7623	0.7283		
Total	35	31.8128			

Figure 22. Plot and Regression Analysis of Normalized Hit Rates and False Alarm Rates for NO-KR Inspectors in Experiment 2



The regression equation is
C33 = - 0.013 + 0.549 C34

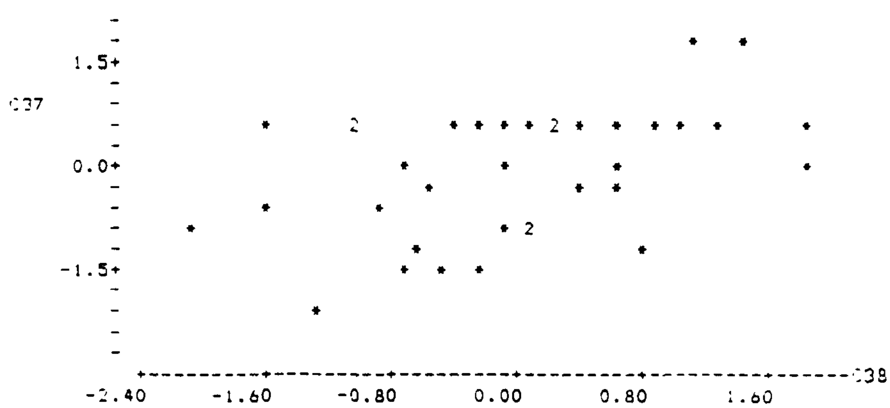
Predictor	Coef	Stdev	t-ratio	p
Constant	-0.0132	0.1328	-0.10	0.922
C34	0.5494	0.1411	3.89	0.000

s = 0.7967   R-sq = 30.8%   R-sq(adj) = 28.8%

Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	1	9.6175	9.6175	15.15	0.000
Error	34	21.5817	0.6348		
Total	35	31.1992			

Figure 23. Plot and Regression Analysis of Normalized Hit Rates and False Alarm Rates for TRUE-KR Inspectors in Experiment 2



The regression equation is
C37 = - 0.021 + 0.411 C38

Predictor	Coef	Stdev	t-ratio	p
Constant	-0.0212	0.1387	-0.15	0.879
C38	0.4111	0.1453	2.83	0.008

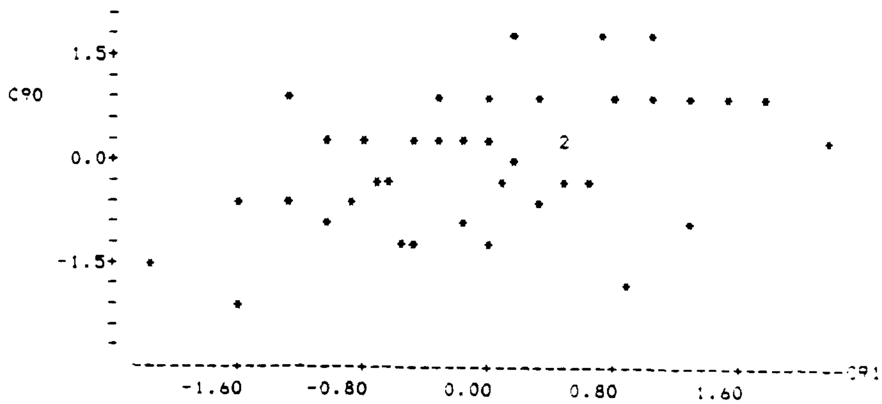
s = 0.8322   R-sq = 19.1%   R-sq(adj) = 16.7%

Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	1	5.5422	5.5422	8.00	0.008
Error	34	23.5490	0.6926		
Total	35	29.0911			

Figure 24. Plot and Regression Analysis of Normalized Hit Rates and False Alarm Rates for FALSE-KR Inspectors in Experiment 2

To further test the assumptions of the SDT model, HR and FAR values in Experiment 3 were converted to z scores and plotted for each Training Group during Phases 2 and 3. Figures 25-28 show that the best fitting lines had positive slopes (less than 1) within each training group. Both experiments satisfied the major assumptions of the SDT model and, therefore, the parameters d' and ß are interpretable for these data.



The regression equation is
C90 = - 0.011 + 0.452 C91

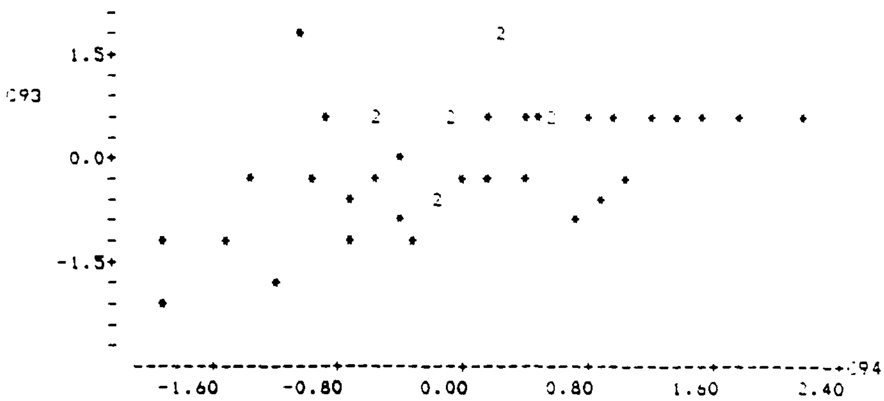
Predictor	Coef	Stdev	t-ratio	p
Constant	-0.0107	0.1349	-0.08	0.937
C91	0.4519	0.1401	3.22	0.003

s = 0.8534 R-sq = 21.5% R-sq(adj) = 19.4%

Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	1	7.5735	7.5735	10.40	0.003
Error	38	27.6775	0.7284		
Total	39	35.2510			

Figure 25. Plot and Regression analysis of Normalized Hit Rates and False Alarm Rates for Group I Inspectors in Experiment 3



The regression equation is
C93 = - 0.022 + 0.447 C94

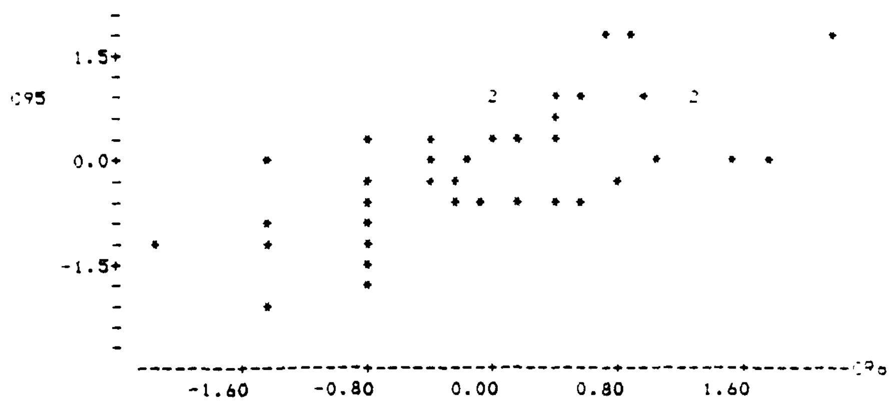
Predictor	Coef	Stdev	t-ratio	p
Constant	-0.0218	0.1289	-0.17	0.867
C94	0.4472	0.1342	3.33	0.002

s = 0.8151 R-sq = 22.6% R-sq(adj) = 20.6%

Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	1	7.3741	7.3741	11.10	0.002
Error	38	25.2473	0.6644		
Total	39	32.6214			

Figure 26. Plot and Regression analysis of Normalized hit Rates and False Alarm Rates for Group II Inspectors in Experiment 3



The regression equation is
C95 = - 0.011 + 0.667 C96

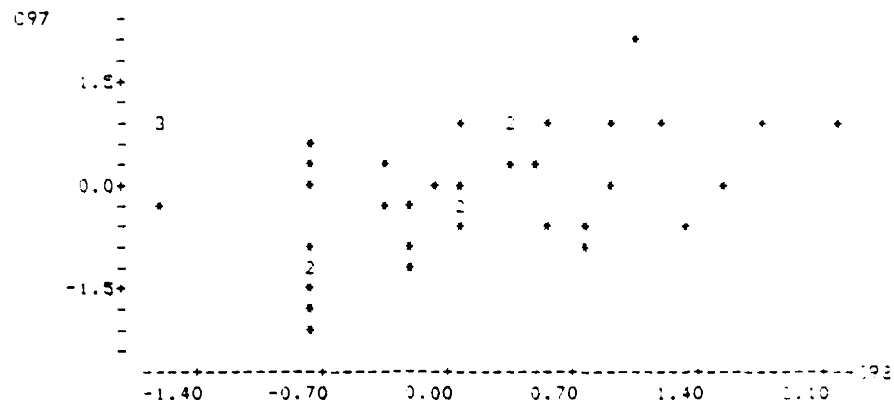
Predictor	Coef	Stdev	t-ratio	p
Constant	-0.0107	0.1131	-0.09	0.925
C96	0.6669	0.1187	5.62	0.000

s = 0.7156   R-sq = 45.4%   R-sq(adj) = 44.0%

Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	1	16.173	16.173	31.59	0.000
Error	38	19.457	0.512		
Total	39	35.630			

Figure 27. Plot and Regression analysis of Normalized hit Rates and False Alarm Rates for Group III Inspectors in Experiment 3



The regression equation is
C97 = - 0.021 + 0.291 C98

Predictor	Coef	Stdev	t-ratio	p
Constant	-0.0208	0.1436	-0.14	0.886
C98	0.2914	0.1532	1.90	0.065

s = 0.9083   R-sq = 8.7%   R-sq(adj) = 6.3%

Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	1	2.9825	2.9825	3.62	0.065
Error	38	31.3490	0.8250		
Total	39	34.3315			

Figure 28. Plot and Regression analysis of Normalized hit Rates and False Alarm Rates for Group IV Inspectors in Experiment 3

VITA

Name: John Micalizzi

Place and date of birth: Brooklyn, N.Y., 7 Nov 1957.

Place and date of high school graduation: Jefferson Township High School, Oak Ridge, N.J. June, 1975.

Places and dates of college graduations:

(1) USAF Academy, Colorado Springs, CO, June 1979, Bachelor of Science.

(2) University of Illinois, Champaign-Urbana, IL, December 1980, Master of Arts, Experimental Psychology.

(3) Penn State University, University Park, PA, May 1990, Ph.D, Industrial Engineering.

Employment History:

(1) July 1979 - Dec 1980. Enrolled in Masters program, University of Illinois.

(2) Jan 1981 - July 1984. Human factors project officer attached to the Ballistic Missile Office, Norton AFB, CA to oversee operational and maintainability design requirements for the Peacekeeper and Minuteman ICBM weapon systems.

(3) Aug 1984 - July 1986. Instructor, Department of Behavioral Sciences and Leadership, USAF Academy, CO.

(4) Aug 1986 - Dec 1989. Enrolled in Ph.D program, Penn State University.

(5) Jan 1990. Associate Professor, Department of Behavioral Sciences and Leadership, USAF Academy, CO.

Publications:

Goldberg, J.H., Micalizzi, J., and O'Rourke, S.A. (1987). The effects of magnification and allowed viewing time on the inspection of printed circuit boards. In Proceedings of the 31st Annual Meeting of the Human Factors Society

Micalizzi, John. (1984). The structure of processing resource demands in monitoring automatic systems. In Proceedings of the 9th Symposium on Psychology in the DOD (USAFA-TR-84-2.; pp. 430-434). USAF Academy, Colorado Springs, CO: Department of Behavioral Sciences and Leadership.

Micalizzi, J., Coward, R.E., and Nelson, W.H. (1979). Simulator training effectiveness as a function of error counts on the F-15A flight simulator instructor-operator station (TR-AFHRL-TR-79-48). Willams AFB, AZ: Flying Training Division.

Micalizzi, J., and Goldberg, J.H. (1989). Decision making in visual inspection: Effects of knowledge of results. In Proceedings of the Industrial Ergonomics and Safety Conference. Cincinnati, OH: University of Cincinnati.

Micalizzi, J., and Goldberg, J.H. (1989). Knowledge of results in visual inspection: Sensitivity or criterion effect? International Journal of Industrial Ergonomics,

Micalizzi, J. and Wickens, C.D. (1980). The application of additive factors methodology to workload assessment in a dynamic system monitoring task (EPL-80-2/ONR-80-2). University of Illinois, IL: Engineering-Psychology Research Laboratory:

Wickens, C.D., Derrick, W.L., Beringer, D.B., and Micalizzi, J. (1980). The structure of processing resources: Implications for task configuration and workload. In Proceedings of the 24th Annual Meeting of the Human Factors Society. Santa Monica, CA: Human Factors Society.

Wickens, C.D., Sandry, D., and Micalizzi, J. (1981). A validation of the spatial variant of the Sternberg Memory Search Task (EPL-81-2/ONR-81-2). University of Illinois, IL: Engineering-Psychology Laboratory.

Professional Organizations:

Air Force Association

Human Factors Society